SKI Report 02:10

Research

The Impact of Spatial Variability of Hydrogeological Parameters - Monte Carlo Calculations Using SITE-94 Data

António Pereira
Robert Broed

March 2002

ISSN 1104-1374
ISRN SKI-R-02/10-SE

SKI perspective

Background

In SKI's deep repository performance assessment project SITE-94 real field data were used on a hypothetical repository for spent nuclear fuel. The project dealt with among other things radionuclide transport calculations in fractured crystalline rock. These calculations were not as comprehensive as was intended from the beginning, especially in the area of probabilistic treatment of hydrogeological parameter uncertainties.

Purpose of the project

The purpose of this project is to elucidate the influence the uncertainties in the hydrogeological parameters have on the radionuclide release to the biosphere. Results from the hydrogeological modelling in SITE-94 are used directly in the one-dimensional radionuclide transport model.

The uncertainties of hydrogeological parameters, Darcy velocity and longitudinal dispersion, together with transport parameters are investigated in Monte Carlo simulations. The Monte Carlo simulations are performed for five radionuclides.

Results

Monte Carlo simulations in which both the hydrogeological and transport parameters are varied simultaneously show considerable spread in the distribution of the peak release rate, compared to when only the hydrogeological parameters are varied. Hence the spread of peak releases is more influenced by the uncertainty of the transport parameters and less by the spatial variability of the hydrogeological parameters. One should, however, bear in mind that there is a synergetic effect between flow and transport phenomena, the impact of which is difficult to disentangle. The number of radionuclides in the investigation is also limited.

From the Monte Carlo simulations one can also conclude that correlations between the two hydrogeological parameters have no appreciable impact on the mean peak value of the maximum releases from the repository.

Effect on SKI's work

This project has clearly shown a useful way to incorporate results from hydrogeological modelling directly into radionuclide transport calculations. This is important knowledge in developing SKI's own capability to perform radionuclide transport calculations.

Project information

Responsible at SKI has been Benny Sundström.

SKI ref.: 14.9-990240/99043.

Relevant SKI report: SKI Report 96:36, SKI SITE-94 Deep Repository Performance Assessment Project, December 1996.

SKI Report 02:10

Research

The Impact of Spatial Variability of Hydrogeological Parameters - Monte Carlo Calculations Using SITE-94 Data

António Pereira
Robert Broed

Department of Physics
Center of Physics, Astronomy and Biotechnology
AlbaNova University Center
Stockholm University
SE-106 91 Sweden

March 2002

SKI Project Number 99043

Contents

Abstract

In this report, several issues related to the probabilistic methodology for performance assessments of repositories for high-level nuclear waste and spent fuel are addressed. Random Monte Carlo sampling is used to make uncertainty analyses for the migration of four nuclides and a decay chain in the geosphere. The nuclides studied are caesium, chlorine, iodine and carbon, and radium from a decay chain.

A procedure is developed to take advantage of the information contained in the hydrogeological data obtained from a three-dimensional discrete fracture model as the input data for one-dimensional transport models for use in Monte Carlo calculations. This procedure retains the original correlations between parameters representing different physical entities, namely, between the groundwater flow rate and the hydrodynamic dispersion in fractured rock, in contrast with the approach commonly used that assumes that all parameters supplied for the Monte Carlo calculations are independent of each other.

A small program is developed to allow the above-mentioned procedure to be used if the available three-dimensional data are scarce for Monte Carlo calculations. The program allows random sampling of data from the 3-D data distribution in the hydrogeological calculations.

The impact of correlations between the groundwater flow and the hydrodynamic dispersion on the uncertainty associated with the output distribution of the radionuclides' peak releases is studied. It is shown that for the SITE-94 data, this impact can be disregarded.

A global sensitivity analysis is also performed on the peak releases of the radionuclides studied. The results of these sensitivity analyses, using several known statistical methods, show discrepancies that are attributed to the limitations of these methods. The reason for the difficulties is to be found in the complexity of the models needed for the predictions of radionuclide migration, models that deliver results covering variation of several orders of magnitude. Correlations between parameters also make it difficult to separate the contribution from each parameter on the output. Finally, it is concluded that even in cases where correlations between parameters can be disregarded for the sake of the uncertainty analysis, they cannot be disregarded in the sensitivity analysis of the results.

A new approach for global sensitivity analysis based on neural networks has been developed and tested on results for the peak releases of caesium. Promising results have been obtained by this method, which is robust and can tackle results from non-linear models even when there are correlations between parameters. This represents a considerable improvement over the capabilities of the commonly used traditional statistical methods.

Abstrakt (Swedish)

I det här arbetet studeras, med tyngdpunkt på den probabilistiska metodologin, flera aspekter relaterade till konsekvensanalyser av ett geologiskt förvar för högaktivt kärnavfall och använt kärnbränsle.

Monte Carlo simuleringar används för att göra osäkerhets- och känslighetsanalyser av radionuklidmigration i geosfären. Nukliderna som studeras är cesium, klor, jod, kol och också radium som resultat av en sönderfallskedja.

En metodologi har utvecklats vars syfte är att överföra tredimensionell data från en hydrogeologisk diskret sprickmodell till endimensionella modeller för transport av radionuklider i berggrunden som används för Monte Carlo beräkningar. Det föreslagna angreppssättet bibehåller de ursprungliga korrelationerna mellan fysikaliska och kemiska fenomen eller processer som t.ex. grundvattenflödet och hydrodynamisk dispersion. Detta innebär en fördel över dagens gängse metod som antar att alla parametrar som ingår i Monte Carlo beräkningar är oberoende från varandra.

Ett dataprogram har utvecklats för att kunna använda den ovannämnda metoden ifall mängden av tredimensionell data är otillräcklig för Monte Carlo beräkningar. Programmet möjliggör sampling av korrelerade slumptal från den empiriska tredimensionella distribution av data som fås från hydrogeologiska beräkningar.

Påverkan av korrelationen mellan grundvattenflödet och hydrodynamisk dispersion på osäkerheten rörande utsläppen av radionukliderna har studerats. För data från SITE-94 är påverkan obetydlig.

Globala känslighetsanalyser från maximumutsläpp av radionukliderna har gjorts. Olika statistiska metoder har använts och resultaten visar vissa diskrepanser som beror på metodernas begränsningar. Orsaken till diskrepanserna ligger i att transportmodellerna är relativt komplexa och dess användning i Monte Carlo beräkningar ger resultat som spänner över flera storleksordningar. Dessutom är det svårt att separera kontributionen av indataparametrarna till slutresultat, när parametrarna är korrelerade. Det konstateras också att man måste ta hänsyn till korrelationer mellan parametrar i en känslighets-analys, även i de fall då korrelationer inte har någon nämnvärd påverkan på sannolikhetsdistribution av radionuklidutsläpp.

En ny metod för känslighetsanalys som använder sig av neurala nätverk har utvecklats och testats på resultat från cesiummigration. De preliminära resultaten pekar på att metoden är robust, kan hantera ickelinjeriteter i utsläppsdistributioner och begränsas inte av korrelationer mellan parametrar något som är i kontrast till de traditionella metoderna för global känslighetsanalys som används i probabilistiska simuleringar. Metoden bör dock utvecklas ytterligare och en flexibel och användarvänlig programvara bör tas fram.

1. Introduction

The SITE-94 project conducted by SKI (Swedish Nuclear Power Inspectorate) is based on an extensive set of field data, part of which is site-specific (provided with courtesy of SKB´s Hard Rock Laboratory at Äspö). The amount of comprehensive data and the results obtained during the exercise provide an important source for further investigations on aspects that were outside the main goals of the SITE-94 project. Such additional work might include probabilistic calculations of the type that may be required in an integrated safety assessment. In this report we use available SITE-94 data to explore some aspects of pertinence to probabilistic calculations. Hence the aim is not to make probabilistic calculations that complement the SITE-94 study, but to use site-specific data in search of less conservative approaches for use within the probabilistic methodology.

We address some aspects that are related to the aleatoric uncertainty and the spatial variability of hydrogeological data. Although spatial variability and aleatoric uncertainty are distinct in origin, it is not possible to disentangle their impact in a straightforward manner. Their synergetic effect can manifest itself in the existence of an underlying correlation between phenomena or data, regardless whether this data is field data or soft data, i.e., data obtained from output from other models.

More specifically, we explore the impact of some input data on uncertainties and how these uncertainties propagate through models of radionuclide transport into the biosphere. The impact is brought about by parameter correlations embedded in some field or soft data, and by the way one commonly uses the data and applies methods in the chain of deterministic or probabilistic calculations linking the source to the biosphere. In this report we focus on these issues in relation to geosphere transport calculations.

We use probabilistic variability calculations to:

- address ways of using hydrogeological data from flow calculations in radionuclide transport calculations
- study the impact of correlated parameters on the uncertainties of radionuclide transport calculations in the geosphere
- examine the performance of global sensitivity analysis for correlated parameters of non-linear models

In the next section we examine some aspects of the data taken from SITE-94 and investigate how it is used in the case variations of the deterministic far-field calculations of that exercise. This is needed for the setting of flow parameters pertinent to our calculations. In Section Three we introduce the cases that will be examined in this report, together with other data used in the models. How we treat the hydrogeological data is explained in Section Four, where the uncertainty and sensitivity calculations are also presented for the case in which only flow parameters are varied. In Section Five uncertainty and sensitivity analyses are presented for cases where the flow and transport parameters are varied simultaneously. The summary and conclusions are included in Section Six followed by the references and appendixes.

2. Premises of the Variability Calculations

2.1 Introduction

SITE-94 terminated with a set of deterministic calculations of radionuclide transport for the near-field and the far-field. Releases to the biosphere were converted to doses. The deterministic calculations were based on data sets resulting from extensive work derived from PID diagram and the FEP methodology[1]. An initially large number of possible combinations of near and far-field calculations were finally reduced to a manageable number of case studies to provide a reference case and a central scenario. Uncertainty and sensitivity analysis was based on the variation calculation results of the deterministic calculations using what-if calculations and variation calculation cases.

As said before, probabilistic calculations were outside the scope of the SITE-94 project. The probabilistic approach has its advantages but also has its inconveniences *vis-à-vis* the deterministic approach. It is not in the scope of this report to discuss this aspect. It is nevertheless clear that a wise combination of the two methods is a convenient way not only to examine the details but also to obtain the synthesis needed in the risk analysis of an integrated performance assessment. Therefore the need to continue exploring certain issues of relevance for the probabilistic methodology.

2.2 Uncertainties in Risk Analysis

In risk assessments one may group uncertainties into two general types, (Helton, 1994): aleatoric (due to the randomness of data) and epistemic (due to incomplete knowledge of the system to be analysed).

Aleatoric uncertainties - also called stochastic uncertainties in some risk literature - express the randomness of data, such as physics and chemical parameters of, for instance, K_d-values. These parameters can be treated statistically by adopting a frequentist approach and, for the purpose of the analysis, are obtained by sampling them from distributions (normal, log-normal, etc.); they are an important part of the input data needed for the mathematical models describing the system.

Whilst expressing our lack of knowledge about the system, for instance of its future long-term evolution, epistemic uncertainties — also sometimes called subjective uncertainties in the literature — are more difficult to treat. An example may be that of the rate number of failing canisters over a long period of time. One way of treating these uncertainties can be to assume a given form for the distribution of that rate of failing canisters based on the elicitation of expert opinion.

Uncertainty and sensitivity analyses are important tools in any integrated performance assessment. In this context, the decision maker may be interested to know which uncertainties can be considered aleatoric or stochastic and which can be viewed as

[1] PID - Process Influence Diagram; FEP - Features Events and Processes

epistemic or subjective. It may not be easy to separate these two kinds of uncertainties (Paté-Cornell, 1996), so it can be argued that the assignment of distributions for parameters like the K_d -values in itself contains a certain degree of subjectivism, leaving the analyst with some mix of stochastic and epistemic uncertainty to deal with. A well founded rational approach to an at least partial degree of objectivism is nevertheless obtained by the systematic use of the FEPs methodology (Feature, Events and Process) and FEPs diagrams which play a very important role in clarifying, systematising and motivating the assumptions embedded in the construction of scenarios or, at a more restricted level, in the grouping of cases for variation calculations.

In any case, for the sake of transparency, it is desirable to separate the two kinds of uncertainties as much as possible. In this report we assume the hydrogeological data that we use, as if its underlying uncertainty is aleatoric. This hydrological data is discussed in the next section. One important aspect of aleatoric uncertainty is the eventual correlation between parameters that will be discussed in Section 2.3.

It is common in probabilistic calculations of radionuclide transport to consider the groundwater velocity and the hydrodynamic dispersion as well as other parameters, as if they belonged to uncorrelated distributions, and consequently to use them as such in the modelling of the system. In deterministic calculations, the relation between those two parameters is uniquely determined because only two single numbers — one for each parameter — enter in the calculation or a functional relation is defined between them. In probabilistic or Monte Carlo calculations, sampling these parameters independently from two distributions may result in unrealistic or improbable combinations of the parameters for groundwater flow velocity and hydrodynamic dispersion.

It can therefore be convenient in Monte Carlo calculations to correlate the two *pdf´s* (probability density function) of those parameters before sampling. However this should not be done in an *ad hoc* manner, but rather, information should be extracted from field data or indirectly from data obtained from hydrogeological calculations of groundwater flow in the modelling domain of interest. The most common way of introducing those correlations is by using the method of rank correlations introduced by Iman and Connover (1982). An alternative method is to use the values of the parameters instead of their ranks as described by Pereira and Sundström (2000). Obviously, if the hydrological data available is sufficient to obtain good statistics from Monte Carlo calculations, one should use it directly. This direct approach is proposed in this report to treat groundwater flow parameters.

2.3 Uncertainty in hydrogeological data

2.3.1 Introduction

Before extracting a data set of flow parameters from SITE-94 for use in the transport calculations, we should take into account some information contained in Table 15.2.12 of the SITE-94 report (1996) which is needed to justify our approach to the study of the propagation of uncertainty in data of flow parameters. That table summarises the combinations of cases for near-field calculations and the far-field variants of that report. The far-field hydrogeological data include data from the so-called Zero Variant (FF0)

and other variations of it. The Darcy velocity for these cases varies by 4.5 orders of magnitude and the longitudinal dispersion by 6.5 orders of magnitude. An interesting feature of that data is shown in Fig. 1: it can be seen that the Darcy velocity and longitudinal dispersion parameters are interrelated from the displacement in the figure. Furthermore, one can see a tendency towards a positive correlation between the Darcy velocity and the dispersion length: high Darcy velocity values are combined with high dispersion lengths and vice versa (note that the axes of the figure are logarithmic). Thus the obvious approach to the choice of the flow parameters for the Monte Carlo calculations is to consider them as belonging to a joint *pdf* so we start our calculations by assuming these two parameters to be an independent pair *vis-à-vis* the remaining flow and transport parameters. At a later stage it will be checked whether it is possible to consider those parameters as being independent for the purpose of the uncertainty and sensitivity analysis. This implies the need to perform calculations in which these parameters are sampled from two distinct *pdfs* instead of using the joint distribution, and the analysis of the associated impact on the uncertainty as well as on the sensitivity of the system to that assumption. (The impact of correlations on both the uncertainty and the sensitivity is considered because they may not influence the uncertainty results, but may still be important for the interpretation of the sensitivity results).

2.3.2 Extracting the flow parameters from the far-field hydrogeological data

We start by examining some of the hydrogeological data used in SITE-94. The values given in Table 15.2.12 of SITE-94 are the result of scrutinising the bulk of the hydrogeological data obtained from the Discrete Feature hydrology model. We will not use the value of the correlation displayed in Figure 1 in our calculations, instead we take some hydrogeological data and use it in Monte Carlo calculations as shown later. The reason is obvious: whenever one has direct access to data, as it is the case here, one should use it.

The two flow parameters – the Darcy velocity and the dispersion length – used in this report are derived from the results of hydrogeological calculations made with the help of a three dimensional discrete feature model for heterogeneous fractured media (Geier, 1996). In a few words, the results of the calculations are obtained by releasing fictive particles at the source, which is situated at a depth of 500 meters, and following their stochastic migration when transported by the groundwater circulating in the main zones and fractures of the crystalline media surrounding the repository. At the surface, one collects data on the positions of the particles, the distance they have travelled and their travel time. By fitting it to the advection-dispersion equation, it is possible to use this data to obtain a Darcy flow and a dispersion length for each particle. (Note that, by themselves, isolated individual particles do not have a dispersion length because a single particle cannot disperse).

Figure 2 shows the two parameters Darcy flow and hydrodynamic dispersion as separate histograms. Figure 3 is a scatter plot that illustrates how the hydrodynamic dispersion (longitudinal dispersion) is related to the groundwater flow rate (Darcy velocity) and Figure 4 is a bivariate histogram of these two parameters, representing their joint *pdf*.

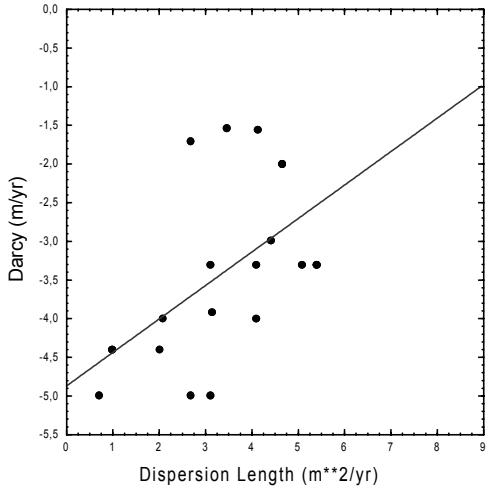

Figure 1 *Plotting the Darcy velocity and the dispersion length data for the cases with different variations used in Site-94 shows the general relationship between these parameters as used from that exercise. The straight line shows the result of a linear regression shown on a log-scale.*

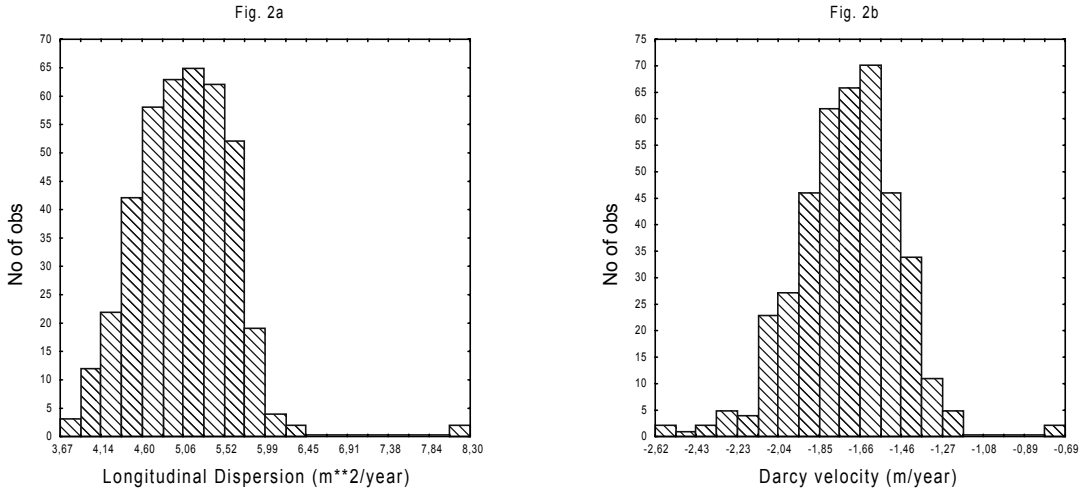

Figure 2 *Histograms of longitudinal dispersion and Darcy velocity obtained by applying the Discrete Feature Model to Site-94 data.*

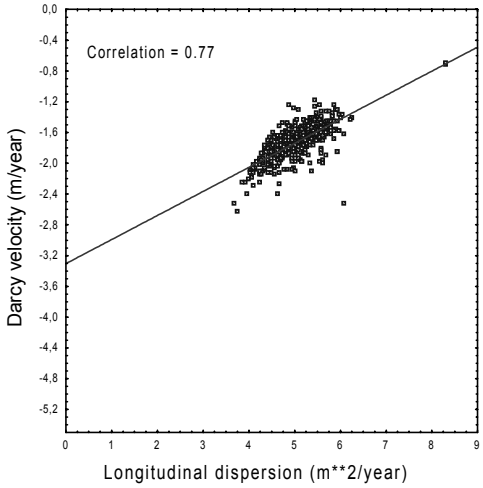

Figure 3 *The relationship between the Darcy velocity and the dispersion length for the data corresponding to the histograms of Figure 2. The straight line shows the result of a linear regression shown on a log-log scale.*

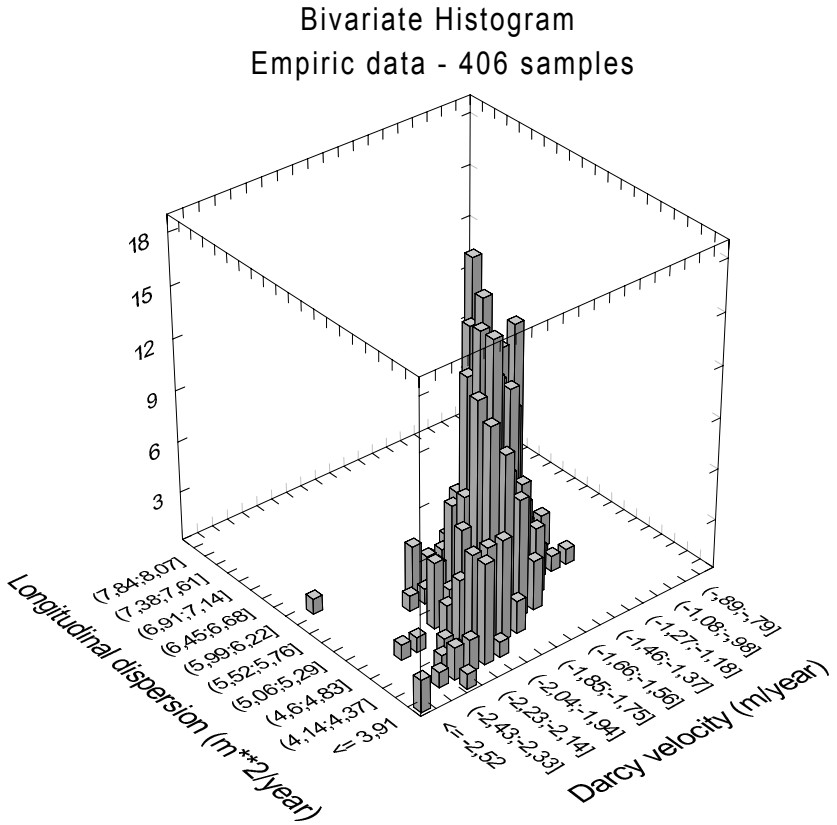

Figure 4 *The bivariate histogram of the Darcy velocity versus the dispersion on a log-log scale, for the data corresponding to the diagram in figure 3.*

Hydrogeological calculations are usually separate from transport calculations in Monte Carlo studies of radionuclide transport in the geosphere because otherwise the computational burden would be overwhelming. The results from hydrogeological simulations are obtained from 2 or 3D modelling, whereas the transport models are, in general, 1D. This poses the question of how to couple the two models, i.e., of how to transfer the data for use in the Monte Carlo calculations of the probabilistic assessment.

In the stochastic calculations, the results of which are used in this work, 406 batches of particles were obtained at the surface, giving 406 pairs of values for the Darcy velocity and longitudinal dispersion. To use this data in the Monte Carlo calculations of radionuclide transport it is usual to take the mean and standard deviations of the values for the Darcy velocities and the longitudinal dispersion (406 values for each in our case) and use them to characterise the central values of the probability distributions for the two parameters. In general, these distributions are assumed to be log-uniform or perhaps uniform. When performing the Monte Carlo calculations, one samples the parameter values for use in the geosphere transport model from these probability distributions, which are *assumed to be independent from each other*.

The above-described indirect method will not be used in this report other than for the purpose of comparing this approach to the direct approach which we claim should be used. The reason is this: in the direct approach we use as much of the information from the hydrogeological calculations as possible and, in particular, we do not disregard the eventual influence of a correlation existing between groundwater flow and dispersion that would be contained in the hydrogeological data. In the direct approach we use the values of the Darcy flow and longitudinal dispersion obtained directly from the discrete feature model, as input data to the geosphere model CRYSTAL (Robinson and Worgan, 1992) embedded in the SYVAC/SU Monte Carlo driver. This is possible because SYVAC/SU allows one to introduce data for the distributions through an input matrix, a feature include by Sundström in the last version of SYVAC/SU [8].

One question that immediately arises is how to obtain adequate statistics for the Monte Carlo calculations. One should consider that although 406 samples may be sufficient for the stochastic calculations made, using the discrete feature model (a balance between needs and computational burden posed by that 3D model) this may be too few samples for the Monte Carlo calculations. In fact, in these calculations one may vary a large number of parameters hence there is a need for a higher number of realisations. In this report we have monitored the convergence of the mean value of the peak release rate independent of time to control the statistics of the calculations. Furthermore, we have also introduced a procedure that obtains an arbitrary number of samples from a fitted distribution to the 3D histogram of Darcy flow versus longitudinal dispersion shown in Figure 4. In this way we also keep the correlation structure between those two parameters. Appendix III includes a listing of the MatLab program developed for this direct approach.

3. Nuclides and input data

3.1 Nuclides included in the calculations

In this report, five nuclides are examined: carbon (^{14}C), chlorine (^{36}Cl), caesium (^{135}Cs), iodine (^{129}I) and radium (^{226}Ra). These nuclides are amongst those that contribute substantially during the first 10 000 years to the intermediate dose potential (IDP) of the near-field for the Zero Variant case (A3) which is the reference case (see the figure on page 554, Vol. II of the SITE-94 report). For the same time period, these same nuclides are also some of the most important contributors to the flux from the far-field for the integrated Zero Variant calculation case in SITE-94 (see the figure on page 587, Vol.II of SITE-94 report). Hence, the choice of these radionuclides for our study.

3.2 Near-field data

The near-field breakthrough curves for each of the nuclides examined are used as the input for the far-field calculations. These curves were taken from the Reference Case, Zero Variant (A3) of SITE-94. For these calculations, it is conservatively assumed that the canister totally fails at a given time point (after 1000 years) , i.e., the whole fuel surface is in immediate and direct contact with the surrounding water from that moment on.

It is most probable that the chlorine, iodine, caesium and radium do not contribute to any solid phase formation. Carbon solubility might eventually be limited by isotope exchange with calcite; nevertheless, solubility limits were considered in the near-field calculations, leading to the source term shown in Figure 5. Radium-226 is a decay product of the U-238 decay chain and exists in the fuel matrix as long as U-238. Its dissolution rate is therefore controlled by the congruent dissolution of the fuel matrix. Reducing conditions are assumed everywhere except on the surface of the fuel.

Observe that for each nuclide we use one and the same near-field breakthrough curve, i.e., no Monte Carlo calculations are made for the near-field. The reason for this is that we are only interested in the implications of uncertainties associated with the hydrogeological, physical and chemical evolution of the far-field.

The near-field breakthrough curves mentioned above are shown in Figure 5. They were obtained from the CALIBRE model (Worgan and Robinson, 1995) (Appendix I). Details of the near-field calculations are not provided here. The interested reader should consult Volume II of the Site-94 report (SKI, 1996) for information on the input data and for the assumptions used in the near-field calculations.

The breakthrough curve for C-14 presented in Figure 5 corresponds to near-field releases from the fuel grain boundaries, the gap release and the release of carbon congruent with matrix dissolution.

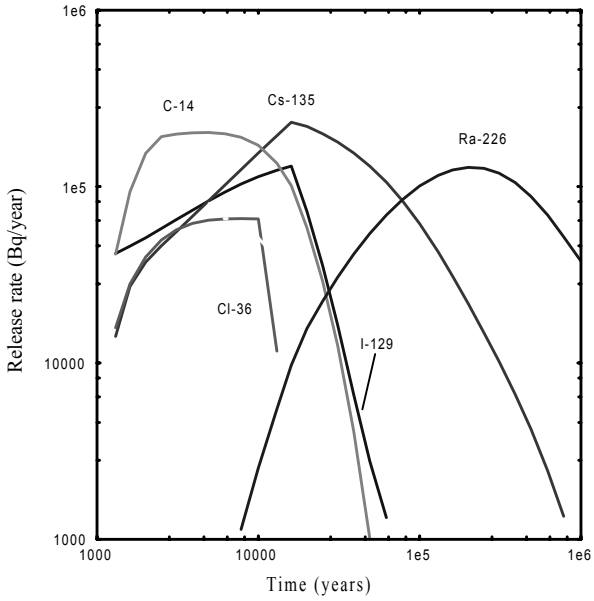

Figure 5 *Near-field breakthrough curves used as the source term for geosphere transport calculations.*

3.3 Far-field data

The Monte Carlo approach is used for the calculations of radionuclide transport in the geosphere. These transport calculations of the selected nuclides are made by keeping some parameters fixed and using probability distributions (*pdf*s) for others. Parameters other than the groundwater velocity and the longitudinal dispersion are sampled when needed using the simple Monte Carlo technique. The calculations take account of the fact that the groundwater flow and the hydrodynamic dispersion are correlated to each other, requiring the use of a joint distribution for these two parameters. Hence the *pdf* used here is a bivariate distribution from which the Darcy velocity and the longitudinal dispersion are sampled. This situation is in contrast to the commonly made assumption in Monte Carlo calculations that these two parameters are independent of each other and, as such, can be sampled from independent probability distributions. The consequences of this situation will be analysed in detail.

To be able to compare the impact of using as much of the information output from the hydrogeological calculations as possible in the Monte Carlo simulations of far-field transport, we need to introduce two concepts that are implicit to the method chosen in this work by asking the following question:

- Is the hydrogeological data *sufficient* for the purpose of the Monte Carlo transport calculations?

At start we have 406 pairs of data points coming from hydrogeological calculations where discrete fracture modelling was used (Geier, 1996). Each pair consists of one Darcy flow parameter and one hydrodynamic dispersion parameter. Now, for Monte Carlo calculations, the 406 pairs of points for the bivariate *pdf* of those two parameters

could be assumed to form a *complete* set of data if one does not intend to vary other parameters in performing the Monte Carlo calculations of radionuclide transport, because, in this case, those points cover the parameter space very well. This can be checked easily by increasing the number of simulations from 406 to, say, 2030, i.e., by a factor of five. If the statistics of our original data set containing 406 pairs of points is complete, the output *pdf* obtained with the increased number of input points should be roughly the same as that resulting from the original 406 data pairs. The question then is how to extract reliable data to extend the number of realisations by a factor of five considering that we have only those 406 data points for the Darcy flow and the hydrodynamic dispersion. We will come back to this point later on.

On the other hand, if we want to vary not only the two flow parameters, but also several other parameters in our Monte Carlo calculations of radionuclide transport, the input parameter space covered by a total of 406 sampling points for each parameter entering in the Monte Carlo calculations may not be *sufficient,* although the information regarding the two water-flow parameters themselves is *complete*. This depends on the number of parameters to be varied simultaneously as can be easily illustrated by the following example: suppose that we want to perform a Monte Carlo calculation where 10 parameters vary simultaneously, two of which are the two key flow parameters, the Darcy velocity and the hydrodynamic dispersion. If each parameter takes only three possible values, the minimum, the median and the maximum value of the *pdf*, then we would need 10^3 realisations instead of 406 to obtain the total number of possible combinations between those 10 parameters, or, for the general case, n^p realisations where n is the number of parameters and p is the number of sampled points per parameter. Looking at our original data given by the histograms displayed in Figure 2, we realise that using three points to characterise each *pdf* is far from enough. At the very least, one should include two more points, the upper and lower quartiles of the distributions, for instance, but this would already necessitate 10^5 samples! In general, however, so many samples are not needed. Nevertheless, to obtain good statistics from all the Monte Carlo calculations with more than two or three variable parameters, it is required that we sample quite a large number of parameter values while retaining the marginal correlations of the original joint *pdf (*or of its histogram). The way we achieve this with the empirical data set of 406 pairs of points as a starting point for the sampling procedure was described earlier.

Case studies

The issue of completeness and sufficiency leads us to the following set of calculations:

- Variation calculations where the flow parameters obtained from the hydrogeology calculations are the only ones to be varied.

- Variation calculations where the flow and transport parameters vary simultaneously

These two sets unfold into the following groups of Monte Carlo calculations:

- Monte Carlo calculations with the original set of flow data, this data is a correlated data set (Sections 4.1 and 5.1)
- Monte Carlo calculations with the original set of flow data, but treating the two parameters as though they were independent of each other (Sections 4.2 and 5.2)

For the second set we have also done Monte Carlo calculations with the number of samples an order of magnitude higher, but using the correlated flow data set (Appendix II).

The flow parameters varying in the Monte Carlo calculations are given directly from the results of hydrogeological modelling using the Discrete Feature Model (Geier, 1996).

4. The impact of the spatial variability of flow parameters

In the probabilistic calculations performed for this work, the random Monte Carlo sampling method is used as the only sampling strategy. Nevertheless, if we are stringent, we should not use the term random Monte Carlo calculations for the set of computations described in Section 4.1 because the only parameters that are varied are the Darcy velocity and the longitudinal dispersion, which are not independent of each other and are not sampled from distributions generated by random Monte Carlo sampling. In fact, they are extracted from an empirical distribution formed by unique and fixed values (not random values) extracted from hydrogeological modelling. It is in the hydrogeological modelling that one uses the field data as an input to calculate the groundwater flow in the geosphere. From the results of this modelling, one extracts the values of the Darcy flow and the hydrodynamic dispersion that form our empirical distribution for use in transport modelling of radionuclides in the far-field. As mentioned before, this two-step approach in which we decouple the flow modelling from the transport modelling is commonly used in probabilistic assessments to reduce the computational burden of the calculations.

4.1 Monte Carlo calculations with correlated input data

The data related to the two flow parameters, the Darcy velocity and the longitudinal dispersion is shown in Table I. This data gives the central values of the marginals of the empirical or joint distribution used in the calculations. In Table II this *pdf* is labelled "empirical". The groundwater flow rate (the Darcy flow) varies by two orders of magnitude and hydrodynamic dispersion by five orders of magnitude. The marginal distribution of the Darcy flow has the shape shown in Figure 2b, while the marginal *pdf* of the hydrodynamic dispersion is shown in Figure 2a.

Table II presents the nuclides studied in this work and a summary of the far-field parameters needed in this section for use in the geosphere transport model CRYSTAL.

Table I *Central values for the marginal distributions of the joint pdf given in Table II.*

Parameter	Darcy flow	Hydrodynamic dispersion
Mean	2.22×10^{-2}	1.18×10^{6}
Confid. (-95%)	2.06×10^{-2}	0.0
Confid. (+95%)	2.38×10^{-2}	2.55×10^{6}
Median	1.99×10^{-2}	1.17×10^{5}
Minimum	2.41×10^{-3}	4.72×10^{3}
Maximum	2.02×10^{-1}	2.00×10^{8}
Lower Quart.	1.35×10^{-2}	4.82×10^{4}
Upper Quart.	2.71×10^{-2}	2.70×10^{5}
Variance	2.62×10^{-4}	1.96×10^{14}
Std. Dev.	1.62×10^{-2}	1.40×10^{7}
Skewness	6.46×10^{0}	1.42×10^{1}

Table II *Far-field parameters values for the constant parameters.*

Parameter	**Element**	**PDF**	**Value**	**Unit**
Darcy velocity, q	All	Empirical	a)	m/year
Longitudinal dispersion, D_L	All	Empirical	a)	m^2/year
Fracture spacing, S	All	Const.	1.0	m
Specific wet. area, a	All	Const.	1.6E-02	m^{-1}
Penetration depth, P_{depth}	All	Const.	5.0E-02	m
Fracture porosity • $_f$	All	Const.	1.0E-03	-
Rock matrix porosity • $_p$	All	Const	5.0E-06	-
Diffusion coefficient in the pore water of the rock matrix, D_m	All	Const.	9.5E-4	m^2/year
Distribution coefficient, K_d	C-14	Const.	1.0E-03	m^3/kg
Distribution coefficient, K_d	Cl-36	Const.	0	m^3/kg
Distribution coefficient, K_d	Cs-135	Const.	1.0E-01	m^3/kg
Distribution coefficient, K_d	I-129	Const.	5.0E-04	m^3/kg
Distribution coefficient, K_d	Cm-246	Const.	5.0E+00	m^3/kg
Distribution coefficient, K_d	Pu-242	Const.	5.0E+00	m^3/kg
Distribution coefficient, K_d	U-238	Const.	5.0E+00	m^3/kg
Distribution coefficient, K_d	U-234	Const.	5.0E+00	m^3/kg
Distribution coefficient, K_d	Th-230	Const.	1.0E+00	m^3/kg
Distribution coefficient, K_d	Ra-226	Const.	5.0E-01	m^3/kg
Migration distance, Z	All	Const.	500.0	m

a) These values are taken from the joint empirical distribution which histogram is shown in Figure 4

4.1.1 Results and Uncertainty Analysis

In this section we present the results of the radionuclide transport calculations and an uncertainty analysis of these results. All results refer to calculations of radionuclide migration escaping from one single canister. Because modelling biosphere transport is not within the scope of this work, the outcomes of the transport calculations refer only to the far-field and therefore the results and the corresponding analysis are expressed in units of Becquerel and not in Sievert.

The measures of importance for the present analysis are based on the release rate of radionuclides independent of time.

Carbon-14 migration

The release rate of C-14 independent of time is shown in Fig. 6. This distribution shows that the uncertainty is small.

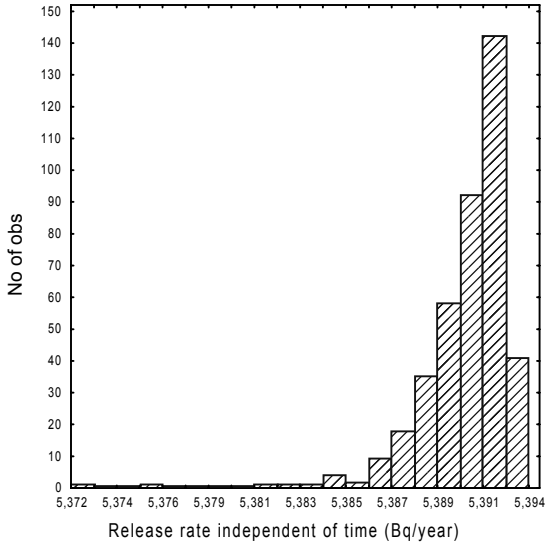

Figure 6 *The far-field histogram for C-14 release. Three decimals are needed in this histogram because of the very low uncertainty of the peak release distribution.*

Caesium-135 migration

The release rate of Cs-135 independent of time is shown in Figure 7. This distribution is left skewed as are the distributions for the remaining nuclides shown in Figures 8 to 10.

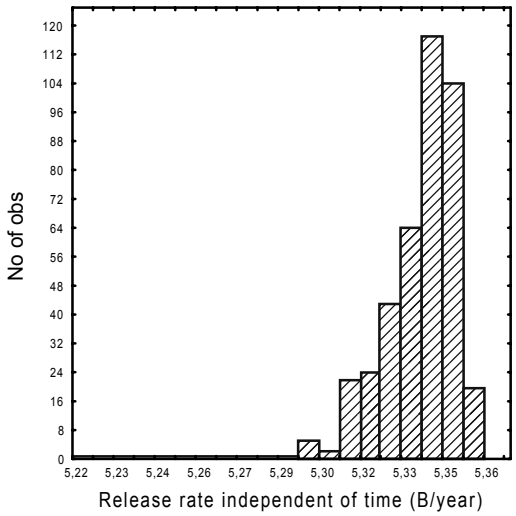

Figure 7 *The far-field histogram for caesium release.*

Cl-36 migration

The release of Cl-36 is in the form of a pulse shape with a lower level tail. This shape results from the fact that Cl-36 is highly mobile with a nil sorption coefficient.

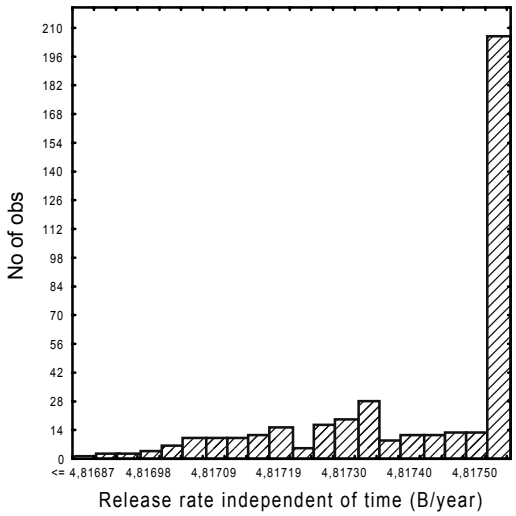

Figure 8 *The far-field histogram for Cl-36.*

I-129 migration

Iodine displays an output distribution that is not as skewed as that of Cl-36. A finite but small distribution coefficient (K_d = 5.0E-04 m^3/kg) was assumed for the sorption of iodine, which is in contrast with chlorine with a nil K_d value.

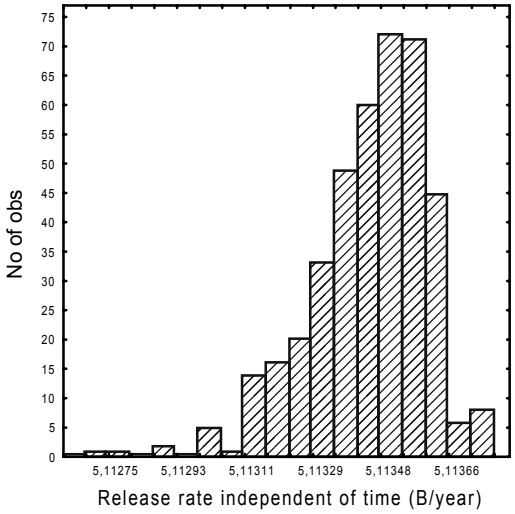

Figure 9 *The far-field histogram for iodine.*

Ra-226 migration

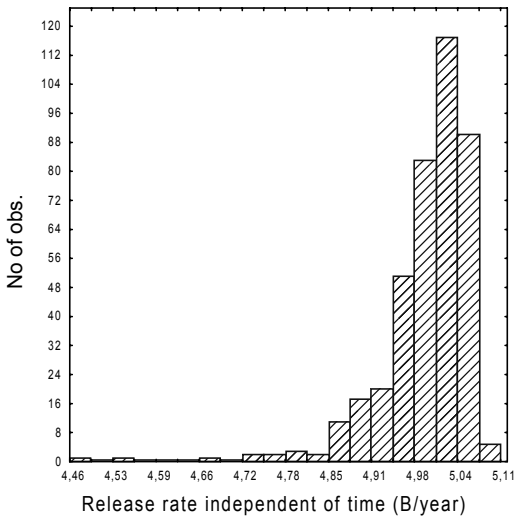

Figure 10 *The far-field histogram for Ra-226 from the decay chain.*

The calculations for all nuclides show an unusually low uncertainty as it is clearly displayed in Table III.

Table III *Central values of peak releases for all nuclides.*

	C-14	**Cs-135**	**Cl-36**	**I-129**	**Ra-226**
Mean	2.458×10^{5}	2.177×10^{5}	6.567×10^{4}	1.298×10^{5}	1.00×10^{5}
Confid. (-95%)	2.457×10^{5}	2.170×10^{5}	6.567×10^{4}	1.298×10^{5}	9.91×10^{4}
Confid. (+95%)	2.460×10^{5}	2.186×10^{5}	6.568×10^{4}	1.298×10^{5}	1.02×10^{5}
Median	2.462×10^{5}	2.193×10^{5}	6.569×10^{4}	1.299×10^{5}	1.03×10^{5}
Lower Quart.	2.454×10^{5}	2.1460×10^{5}	6.566×10^{4}	1.298×10^{5}	9.39×10^{4}
Upper Quart.	2.466×10^{5}	2.226×10^{5}	6.570×10^{4}	1.299×10^{5}	1.10×10^{5}
Std. Dev.	1.20×10^{3}	6.93×10^{4}	2.73×10^{1}	4.95×10^{1}	6.77×10^{2}

The parameters that are varied in the simulations are the two flow parameters Darcy flow and hydrodynamic dispersion. The Darcy flow has an uncertainty of two orders of magnitude and the hydrodynamic dispersion varies by almost five orders of magnitude. Yet the propagation of these uncertainties does not result in a wide uncertainty in the maximum release rate independent of time. The explanation for this observation is given by the sensitivity analysis done in Section 5.

The mean value of the maximum release rate (peak release) for the Monte Carlo calculations of all nuclides using 406 samples is expected to have converged because we vary only two parameters. A curve illustrating the mean peak as a function of the number of samples is shown in Figure 11 for caesium. A sudden jump is observed at around 200 samples, it is caused by the inclusion of an “extreme” parameter value when the number of samples goes from 200 to 250 samples. In fact it is a small percentage increase. When we increase the number of variable parameters by including random variations of other flow parameters and of transport parameters, the convergence of the Monte Carlo calculations should be carefully considered.

4.1.2 Sensitivity Analysis

As the calculations performed in this section only deal with two variable parameters, the Darcy velocity and the longitudinal dispersion, one could anticipate a straightforward sensitivity analysis. Unfortunately this is not the case, which has been the motivation for pursuing this analysis.

We have applied two different statistical approaches to global sensitivity analysis: Spearman statistics and the Pearsson correlation coefficients approach. The first approach is a method of non-parametric statistics based on ranks whist the second is based on raw data values and not on the ranking of those values. The analysis is done for the peak releases, i.e., the maximum release rates independent of the time at which they occur.

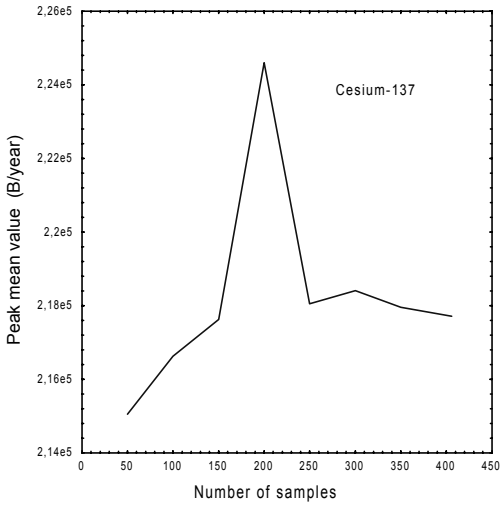

Figure 11 *The variation in the peak release of ^{135}Cs with number of samples.*

The results computed with the help of the approaches mentioned above are shown in Table IV. Both methods identify the Darcy velocity as the most important parameter. Qualitatively they are also consistent in the sense that in each case an increase in the Darcy velocity implies an increase in the peak release rate; the same prediction is made for increasing values of the longitudinal dispersion. The results of the Spearman test for caesium and iodine are consistent with other non-parametric tests (Kendall and Gamma tests), the results of which we do not include here. For all non-parametric tests the numerical values are relatively close: the Darcy velocity shows a correlation with the peak release which varies between 0.91 and 0.98 and the longitudinal dispersion coefficient varies between 0.60 and 0.79. The Spearman test is "less" non-parametric than the Kendall and Gamma tests.

Table IV *Sensitivities of peak release for the two correlated flow parameters.*

Correlation coefficients			
Nuclide	**Variable**	**Peak**	
		Pearsson	Spearman
Cs-135	Darcy velocity	0.60	0.98
	Long. Dispersion	0.13	0.79
I-129	Darcy velocity	0.66	0.98
	Long. Dispersion	0.15	0.78
Cl-36	Darcy velocity	**-0.02**	**-0.14**
	Long. Dispersion	**+0.06**	**-0.12**
Ra-226	Darcy velocity	0.64	0.98
	Long. Dispersion	0.15	0.80
C-14	Darcy velocity	0.55	0.98
	Long. Dispersion	0.77	0.78

The quantitative predictions vary strongly between the Pearsson correlation and the Spearman test. Observe however that one cannot expect the same absolute numerical values from different tests, so when we talk about a strong variation in the results, we refer to the relative difference between the numerical values for the two parameters within each test. For example, for Cs-135 the Pearsson test puts the Darcy velocity as the first parameter, it also tells us that it is much more important than the longitudinal dispersion (0.60 against 0.13 respectively). Recalling our previous knowledge that these two parameters, which come from hydrogeological information, are strongly correlated (linear correlation of 0.71), it seems clear that one or both methods break down in the presence of correlations between the *pdf* of the parameters. Figure 12 shows clearly the impact of these correlations on the peak release. Any horizontal slice at any Z=constant value, displays the elliptical correlation pattern shown in the scatter plot for these two parameters in Figure 3. The most striking result is the complete breakdown of the two statistical methods for the results of the Cl-36 migration. The reason is the very small uncertainty associated with the strong low release tail of the distribution for Cl-36 (Figure 8).

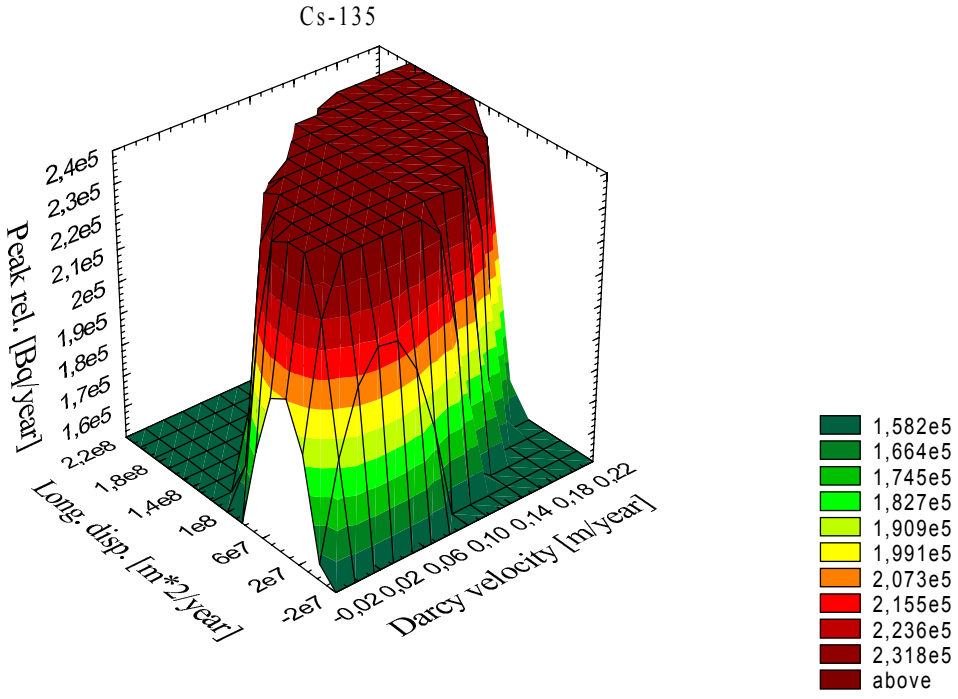

Figure 12 *The relation between the variable parameters and the peak release of Cs-135. The graph shows clearly the correlation between both parameters and the impact of these parameters on the release.*

It is well known that the global sensitivity methods do not perform well for non-linear models in the presence of correlations between parameters (Prado et. al., 1999). The differences apparent in Table IV are indicative of the strong influence of the correlation

between the two flow parameters. It is difficult at present to know which method best represents the truth.

The methods described above are correlation methods. Recently developed sensitivity methods based on variance, for instance the Sobol indices (Sobol, 1993) and Fourier Amplitude Sensitivity (Saltelli and Bolado, 1998), can compute higher-order terms for sensitivity indices and total indices whenever the results are non-linear functions of the input parameters, but unfortunately they cannot handle the case of correlated parameters. In summary, even when applied to the apparently simple case of two varying parameters that are correlated, the methods used in this section and other methods can only give us qualitative results.

4.2 Monte Carlo calculations with uncorrelated input data

The approach suggested in this report to reduce the hydrogeological data to the format needed by the transport models conserves the correlation structure of the data. In the calculations made in this report only two parameters appear as correlated data: the Darcy flow and the longitudinal dispersion. The immediate question is therefore: what is the impact of this correlation on the results of the transport models.

The straightforward way to answer to this question is to "decorrelate" those parameters and repeat the calculations in order to compare with the results obtained in Section 4.1. This "decorrelation" is made by random mixing of the data. If *X1(N)* is a vector containing the Darcy parameter values and *X2(N)* the one containing the longitudinal dispersion values (with N=406 realisations in our case), we can take each of them separately and perform a random permutation of their values. After this operation the correlation values between the Darcy and the longitudinal dispersion values is nil.

4.2.1 Results and Uncertainty Analysis

The scatter plot of Figure 13 shows the flow field data used in the Monte Carlo calculations. In this figure one can see a region with a structure that is non-random. The original pattern obtained by the permutation is perfectly random, but some of the pairs of values for the Darcy velocity q, and the longitudinal dispersion D_L result in Peclet numbers higher than 100 which cannot be handled by the CRYSTAL model used in the transport calculations. The Peclet number is defined by:

$$Pe = qL/\theta D_L$$

where q [m/year] is the Darcy velocity, L [m] is the transport length, • [-] is the flow porosity and D_L [m^2/year] is the longitudinal dispersion coefficient. Physically the Peclet number is the ratio of the dispersive over the convective (or advective) term in the transport equation.

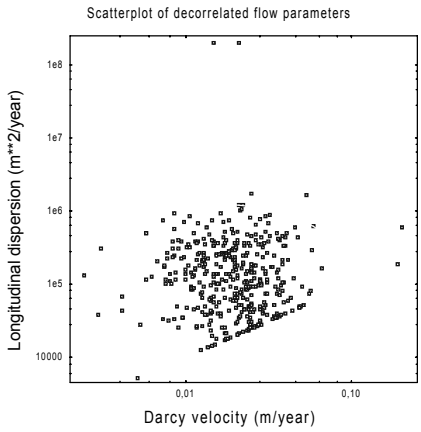

Figure 13 *Scatter plot of Darcy velocity versus longitudinal dispersion parameters. The graph shows that the parameters are "decorrelated". The special linear structure at the bottom of the picture results from resetting with unphysical pairs of data (see text).*

What has been done to tackle this situation was to reset the longitudinal dispersion values leading to Peclet numbers that are greater than 100 to the highest value compatible with a Peclet number equal to 100. In this way one can avoid physically unreasonable combinations of parameters. Although some combinations of Darcy values with longitudinal dispersion values such that the Peclet is higher than 100 could be physically reasonable, the difference in the output consequences is insignificant.

In this section an uncertainty analysis is done with the same input data for the Monte Carlo calculations as that used in Section 4.1.1 with the exception of the hydrogeological parameters Darcy flow and longitudinal dispersion, which are those obtained from the original data set after decorrelation.

The results obtained with this new input data are presented in Table V and as histograms (Figure 14). Comparing these results with those of Table III for the case of correlated flow data, it is observed that the correlation has no impact on the mean values of the peak release rates for any nuclide. This is not always the case. For instance, Pereira (Pereira and Sundström, 2000) found a discrepancy of half an order of magnitude for the mean peak release in Monte Carlo calculations between correlated and non-correlated data.

The results obtained in Table V are as expected. In fact, in the situation modelled here the geosphere has an extremely low impact on the attenuation of the mean peak releases as is clearly illustrated by Figure 15. In this figure, we have plotted the near-field curve of Cs-135 together with the ten breakthrough curves, which have the highest peak releases for the simulated nuclide. These ten curves dominate over the mean peak values. We see that in every case considered here the peak of the far-field curves is almost equal to the respective near-field breakthrough curve.

In summary, for the input data used in these calculations, the correlation between the Darcy velocity and the longitudinal dispersion is of no importance for the peak release *pdf,* i.e., for release rates independent of the time of occurrence. Therefore it is also

expected that the sensitivity indices obtained in the sensitivity analyses of the correlated and uncorrelated data will be almost the same (see next section). The impact of the correlations on the *pdfs* at different points in time will be examined later on.

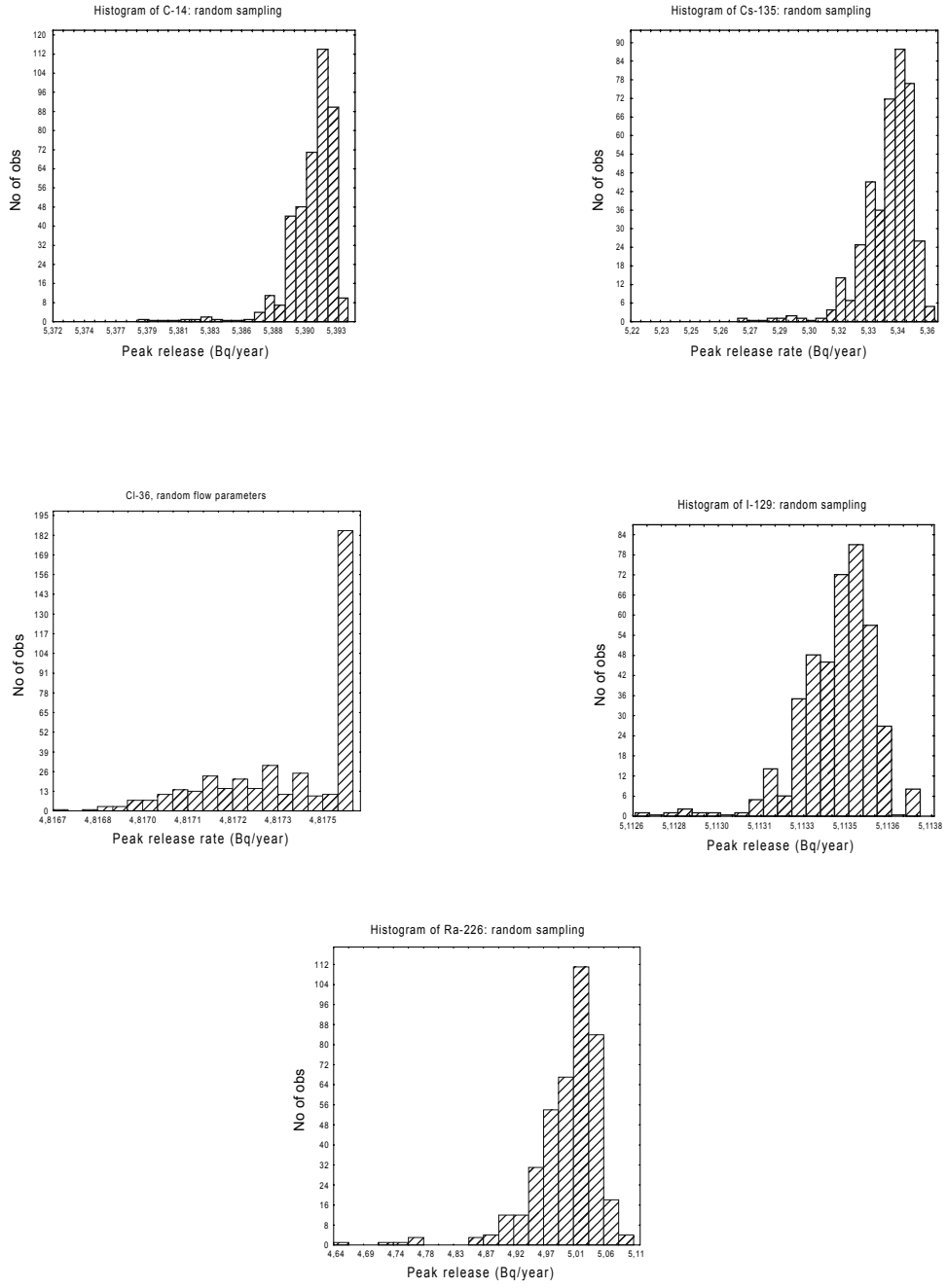

Figure 14 *The histograms of peak release rates for random sampling of the flow parameters.*

Table V *Central values for all nuclides for uncorrelated parameters.*

	C-14	**Cs-135**	**Cl-36**	**I-129**	**Ra-226**
Mean	2.460×10^{5}	2.185×10^{5}	6.570×10^{4}	1.300×10^{5}	1.021×10^{5}
Confid. (-95%)	2.459×10^{5}	2.179×10^{5}	6.568×10^{4}	1.298×10^{5}	1.010×10^{5}
Confid. (+95%)	2.461×10^{5}	2.190×10^{5}	6.568×10^{4}	1.299×10^{5}	1.032×10^{5}
Median	2.462×10^{5}	2.197×10^{5}	6.565×10^{4}	1.299×10^{5}	1.042×10^{5}
Lower Quart.	2.456×10^{5}	2.152×10^{5}	6.570×10^{4}	1.298×10^{5}	9.604×10^{5}
Upper Quart.	2.466×10^{5}	2.223×10^{5}	6.570×10^{4}	1.299×10^{5}	1.094×10^{5}
Std. Dev.	9.560×10^{2}	5.680×10^{3}	3.040×10^{1}	4.280×10^{1}	1.130×10^{4}

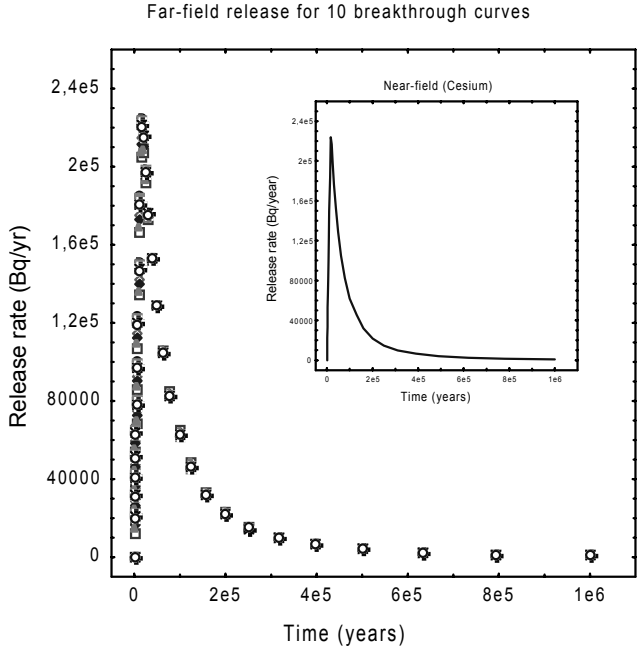

Figure 15 *The relation between near-field and some far-field breakthrough curves for Cs-135.*

4.2.2 Sensitivity Analysis

The sensitivity analysis conducted here uses three non-parametric methods and one parametric one (the Pearsson correlation coefficients). The results are summarised in Table VI.

The sensitivity results shown in this table are somewhat different from those shown in Table IV for the case of the Spearman coefficients. The difference found here for the coefficients of Darcy flow and longitudinal dispersion using the two different statistical

methods is approximately the same, which was not the case for the calculations performed with correlated parameters.

Table VI *Sensitivity values for uncorrelated data.*

Correlation coefficients for uncorrelated flow parameters

Nuclide	Variable	Peak			
		Pearsson	Spearman	Kendal	Gamma
Cs-135	Darcy velocity	0.61	0.94	0.83	0.83
	Long. Dispersion	0.14	0.22	0.15	0.15
I-129	Darcy velocity	0.65	0.94	0.83	0.84
	Long. Dispersion	0.16	0.22	0.15	0.14
C-14	Darcy velocity	0.55	0.94	0.83	0.83
	Long. Dispersion	0.11	0.21	0.15	0.15
Cl-36	Darcy velocity	**0.04**	**-0.07**	**0.0**	**-0.01**
	Long. Dispersion	**0.06**	**-0.07**	**-0.05**	**-0.06**
Ra-226	Darcy velocity	0.62	0.93	0.81	0.81
	Long. Dispersion	0.17	0.25	0.17	0.17

This indicates that the rank method is somewhat more sensitive to the presence of correlations as shown in Table IV.

Furthermore for these Monte Carlo calculations with uncorrelated parameters, the sensitivity analysis methods break down for the Cl-36 radionuclide migration. This result was also found for similar calculations in which the parameters were correlated; hence, the fact that the parameters are correlated is not the explanation for the failure.

5. The impact of flow and transport parameter uncertainties

5.1 Monte Carlo calculations with correlated input data

In Section 4 it was demonstrated that although input data varied by at least two orders of magnitude, the two flow parameters, Darcy velocity and longitudinal dispersion, did not have an impact on the uncertainty in the peak release rates, the *pdf* of this distribution having a very low standard deviation. The explanation for this is that the transport parameters were kept constant during the Monte Carlo simulations.

In this section we organise the calculations in the following way: first, we make some exploratory calculations on Cs-135 in which we systematically increase the number of variable parameters from calculation to calculation. The results are shown in Appendix II. Second, we make radionuclide transport calculations for all nuclides, but now varying all parameters simultaneously.

The goal of the first set of exploratory calculations was to illustrate how the uncertainty in the flow and transport parameters propagates through the models for each new parameter that we allow to vary. These calculations resulted in a considerable broadening of the distribution representing the consequences that would be observed. Hence, the small uncertainty observed originally in the propagation of the two hydrogeological parameters, the advective velocity and the longitudinal dispersion is masked in the final results of these calculations.

The goal of the second set of calculations was to examine the impact of the uncertainty in the parameters for all nuclides considering that the hydrogeological data representing the Darcy velocity and the longitudinal dispersion are correlated.

5.1.1 Results and uncertainty analysis

In the set of calculations used in this section, the parameters that are considered to be uncertain for each nuclide are varied simultaneously. The input data for all parameters is displayed in Table VII. The variable parameters are sampled within certain intervals taken from different sources. The boundaries of the distribution coefficients come from: Cl-36 – lower limit (Table 15.2.7), Site-94 report and upper limit, TVO-92 report; C-14 – lower limit, Site-94 report and upper limit data from fracture fillings SKI-TR: 96-2 pp. 7; Cs-135 – lower limit Site-94 report and upper limit chosen as a factor of the lower limit; I – lower limit Site-94 report and upper limit 2.3 times higher than low limit (less than for the oxidising case); Cm-246 – lower limit from Site-94 report and upper limit equal to the value of far-field rock given by SKI-TR:96-2, pp. 22; Pu-242 – lower limit from fracture filling III reducing conditions SKI-TR:96-2, pp.21 and upper limit from Site-94; U234 and 238 – lower limit from fracture filling III reducing conditions SKI-TR:96-2, pp.19 and upper limit from SITE-94 report; Th-230 – lower limit from fracture filling III reducing conditions, SKI-TR:96, pp.17 and upper limit from Site-94; Ra-226 – lower limit from TVO-92 report and upper limit from SITE-94 report.

The results are summarised in Table VIII and by the histograms of Figure 16. The probabilistic results of the uncertainty analysis are presented in Fig 17 where they are given by the mean value of the peak releases and confidence levels of ±95%. In the logarithmic scale, the confidence intervals lie very close to the respective mean values. The deterministic results of SITE-94 data (Table 16.3.4, Vol. II) are shown in the same figure. One can observe that the deterministic maximum releases of C-14 and I-129 according to the data from SITE-94 are near the mean values of the peak releases of the probabilistic calculations. For Cl-36 the deterministic results differ by one order of magnitude and for Cs-135 and Ra-226, the discrepancy is higher than one and two orders of magnitude, respectively. But one should consider that the parameter variations in the probabilistic calculations not only cover the integrated "Zero Variant" case for the "Reference Case" of the SITE-94 study, but also other more conservative cases given in the SITE-94 report.

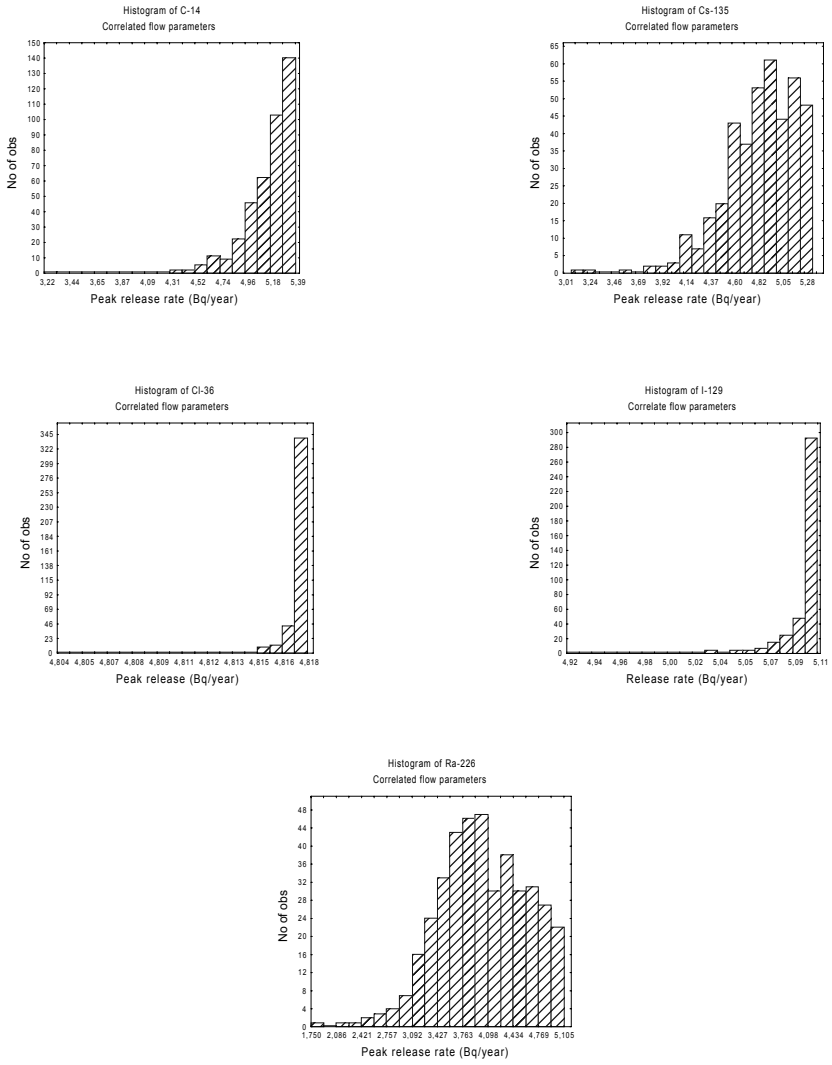

Figure 16 *– Histograms of the peak release rates for the case of correlated parameters.*

Table VII *Parameters and distributions used in the calculations.*

Parameter	Element	PDF	Value	Unit
Darcy velocity, q	All	Empirical	a)	m/year
Longitudinal dispersion, D_L	All	Empirical	a)	m^2/year
Fracture spacing, S	All	Const.	1.0	m
Specific wet. area, a	All	Uniform	6.3E-03 7.86E-01	1/m
Penetration depth, P_{depth}	All	Uniform	2.0E-02 2.50E-01	m
Fracture porosity • $_f$	All	Const.	1.0E-03	-
Rock matrix porosity • $_p$	All	Const	5.0E-06	-
Diffusion coefficient in the pore water of the rock matrix, D_m	All	Const.	9.5E-4	m^2/year
Distribution coefficient, K_d	C-14	Uniform.	1.0D-3 1.0D-2	m^3/kg
Distribution coefficient, K_d	Cl-36	Uniform	0.0 1.0D-4	m^3/kg
Distribution coefficient, K_d	Cs-135	Uniform	1.0E-01 5.0E-01	m^3/kg
Distribution coefficient, K_d	I-129	Uniform	3.0E-04 7.0E-04	m^3/kg
Distribution coefficient, K_d	Cm-246	Uniform	0.0 5.0	m^3/kg
Distribution coefficient, K_d	Pu-242	Uniform	2.0 5.0	m^3/kg
Distribution coefficient, K_d	U-238	Uniform	1.0 5.0	m^3/kg
Distribution coefficient, K_d	U-234	Uniform	1.0 5.0	m^3/kg
Distribution coefficient, K_d	Th-230	Uniform	0.1 1.0	m^3/kg
Distribution coefficient, K_d	Ra-226	Uniform	0.2 0.5	m^3/kg
Migration distance, Z	All	Const.	500.0	m

a) data given by the empirical distribution (joint *pdf*) shown in Fig. 4.

Table VIII *Central values for all nuclides*[†].

	C-14	Cs-135	Cl-36	I-129	Ra-226
Mean	1.583×10^5	9.406×10^4	6.562×10^4	1.267×10^5	2.271×10^4
Confid. (-95%)	1.527×10^5	8.850×10^4	6.561×10^4	1.261×10^5	2.001×10^4
Confid. (+95%)	1.639×10^5	9.961×10^4	6.563×10^4	1.272×10^5	2.542×10^4
Median	1.617×10^5	8.259×10^4	6.563×10^4	1.286×10^5	1.041×10^4
Lower Quart.	1.172×10^5	4.563×10^4	6.561×10^4	1.266×10^5	4.273×10^3
Upper Quart.	2.096×10^5	1.362×10^5	6.566×10^4	1.295×10^5	2.949×10^4
Std. Dev.	5.735×10^4	2.825×10^3	1.185×10^2	5.759×10^3	2.776×10^4

[†] Three decimals are used in this table because the small uncertainty in Cl-36 results.

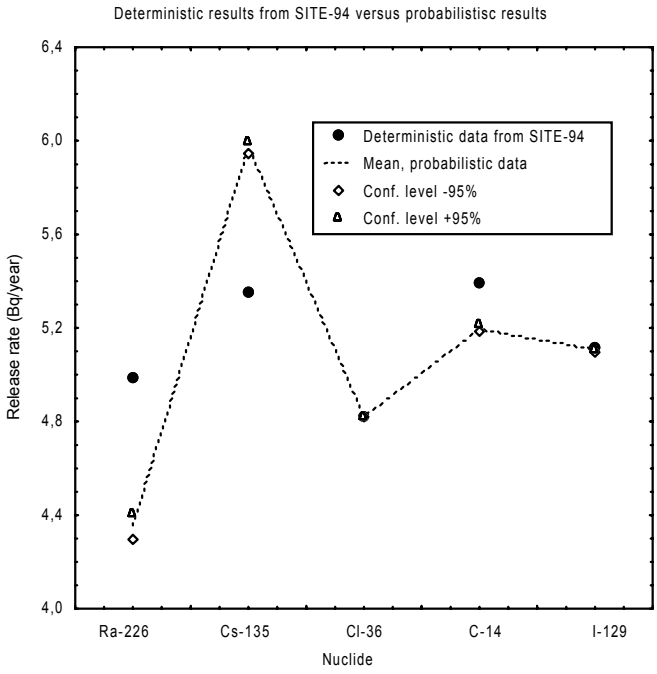

Figure 17 *Deterministic results from SITE-94 (low F-ratio variation case) versus probabilistic results from this work.*

5.1.2 Sensitivity analysis

Statistical methods of global sensitivity analysis

In this section we perform the sensitivity analysis for output distributions in calculations in which five parameters were simultaneously varied: the Darcy velocity q, the hydrodynamic or longitudinal dispersion D_L, the flow wetted surface area a, the penetration depth P_{depth} and the distribution coefficient K_d. One should bear in mind**,** that in general a sensitivity analysis is not as dependent on the number of samples as an uncertainty analysis. Obviously, if we have a relatively low number of simulations, the numerical results will not be as exact as if the convergence of the Monte Carlo calculations had been fully attained by using a very high number of samples, but we can still get the general trends about the relative importance of the different parameters if we are not far away from a full convergence of the Monte Carlo calculations. It is the case here, as is shown by the rapidly decreasing oscillatory behaviour of the mean peak shown in Figure 18.

In this section we again use two methods of sensitivity analysis, the first based on the raw values for the peak releases and the other using methods of non-parametric statistics based on ranks of those peaks. The results of the sensitivity analysis are summarised in Table IX.

For each nuclide we observe a general consistency between the results of the non-parametric test (Spearman) and the Pearsson correlation coefficient with one important exception: for all nuclides the Spearman correlation coefficient for the longitudinal dispersion parameter is very different from the Pearsson correlation coefficient. It is most likely that this disagreement arises because the Darcy flow is correlated to the longitudinal dispersion. It is not possible from the results to explain why the difference appears in the longitudinal dispersion parameter and not in the Darcy correlation parameter; in this respect there is also one exception, the Darcy correlation parameter for the iodine nuclide from the Pearsson statistics is quite different from that from the Spearman statistics.

Another interesting observation is related to the values of Cl-36 given by both statistics, which are generally lower for all parameters than the values for the other nuclides. This is probably caused by to the very low variance of the peak release associated with the skewness of its distribution.

Resuming our discussion of the results displayed in Table IX below, we observe that the wetted surface area a is the most important parameter for all nuclides apart from Cl-36 (for Pearsson statistics). The second most important parameter is the groundwater flow, the exceptions being for Cl-36 and I-129 (Pearsson statistics); in general the distribution coefficient is the third most important parameter according to Pearsson statistics whereas this place is taken by the longitudinal dispersion according to the Spearman statistics. The penetration depth is important only for Cl-36 and I-129, which are the nuclides with very low distribution coefficients. This indicates that if these nuclides are retarded, it is because of the diffusion mechanism in the matrix, but without retention due to a K_d -effect.

Cl-36 is the nuclide with lowest dispersion and that, which poses difficulties when subjecting the two sets of statistics to the sensitivity analysis. We have seen before that global sensitivity analysis breaks down totally for this nuclide when only the flow parameters are varying (Table IV).

Table IX *Results of the sensitivity analysis with two correlated flow parameters. The integers in bold represent the parameter rank.*

	Cs-135				**C-14**				**Ra-226**			
	Peak release				Peak release.				Peak release.			
Parameter	Pearsson		Spearman		Pearsson		Spearman		Pearsson		Spearman	
Q [m/year]	0.43	**2**	0.55	**2**	0.41	**2**	0.52	**2**	0.45	**2**	0.55	**3**
K_d [m^3/kg]	-0.18	**3**	-0.20	**4**	-0.27	**3**	-0.27	**4**	-0.10	**4**	-0.18	**4**
D_L [m^2/year]	0.17	**4**	0.46	**3**	0.11	**4**	0.44	**3**	0.27	**3**	0.57	**2**
P_{depth} [m]	-0.001	**5**	-0.005	**5**	-0.06	**5**	-0.06	**5**	-0.03	**5**	-0.008	**5**
a [1/m]	-0.80	**1**	-0.80	**1**	-0.74	**1**	-0.77	**1**	-0.71	**1**	-0.73	**1**

	Cl-36				**I-129**			
	Peak release				Peak release			
Parameter	Pearsson		Spearman		Pearsson		Spearman	
Q [m/year]	0.15	**4**	0.16	**4**	0.29	**3**	0.46	**2**
K_d [m^3/kg]	-0.22	**2**	-0.20	**3**	-0.17	**4**	-0.18	**5**
D_L [m^2/year]	0.05	**5**	0.16	**4**	0.04	**5**	0.39	**3**
P_{depth} [m]	-0.24	**1**	-0.21	**2**	-0.33	**2**	-0.37	**4**
a [1/m]	-0.20	**3**	-0.31	**1**	-0.39	**1**	-0.71	**1**

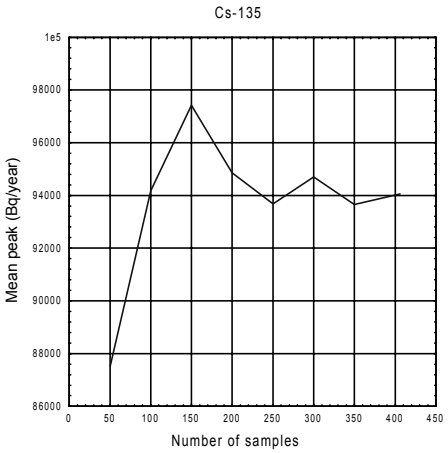

Figure 18 *The variation of the mean peak value of ^{135}Cs with increasing number of samples. Five parameters are varied in this calculation. The oscillatory behaviour indicates that that the calculation is approaching convergence.*

Finally we observe that in all cases there is an agreement on the type of impact that the different parameters of the input distributions have on the output distribution: the Darcy flow is positive, i.e., the higher this parameter is the higher is the peak release rate. The distribution coefficient K_d has a negative correlation showing that the lower the K_d is, the higher the peak release rate will be. The correlation value signals that the penetration depth, P_{depth} and the wetted surface are also in accordance with the physical meanings of the parameters, although the two statistics do not work properly in the presence of the correlation between groundwater flow (Darcy velocity) and longitudinal dispersion.

The sensitivity of the consequences to the variation of the parameters can be visualised by "sensitivity plots" (figure 18). Qualitatively one can say that the greater the distance of the cumulative curve to the diagonal, the more sensitive is the parameter.

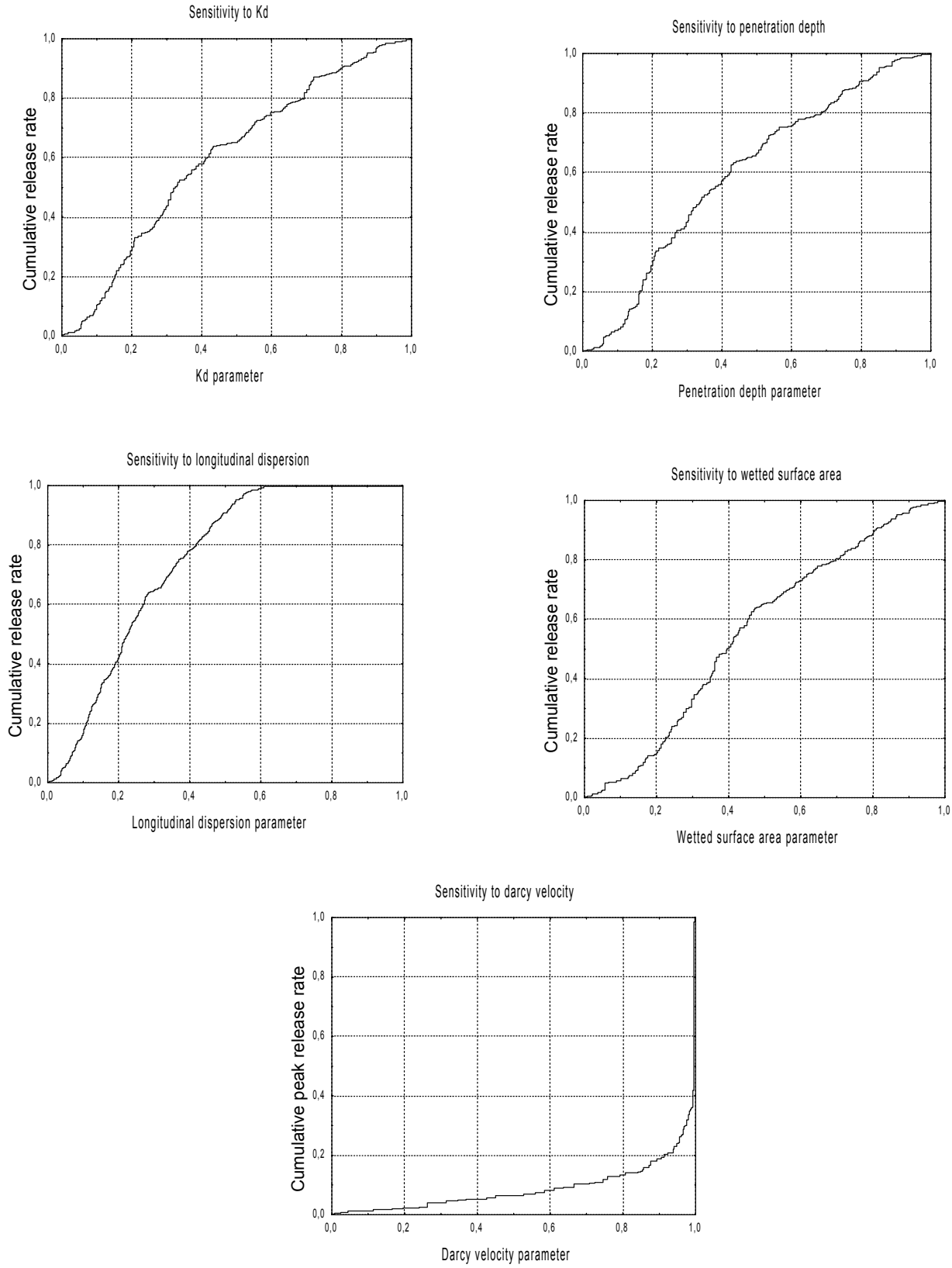

Figure 19 *The sensitivity plots of Cs-135 illustrate the relative importance of the parameters.*

5.2 Monte Carlo calculations with uncorrelated input data

5.2.1 Results and Uncertainty Analysis

The results of the calculations for uncorrelated flow data are shown as histograms of peak release rates on the left hand side of Figures 20 to 24. On the right we have included the results obtained in Section 5.1.1 to make comparison between the respective plots easier.

Cl-36

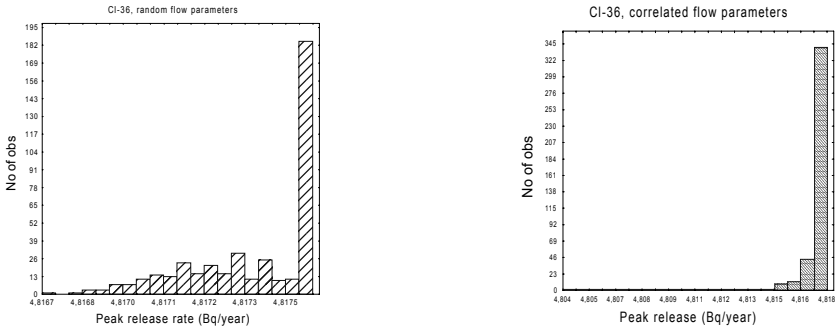

Figure 20 *The release rate histogram for Cl-36. The picture to the left corresponds to random sampling of flow parameters and that to the right to the case of correlated flow parameters.*

Ra-226

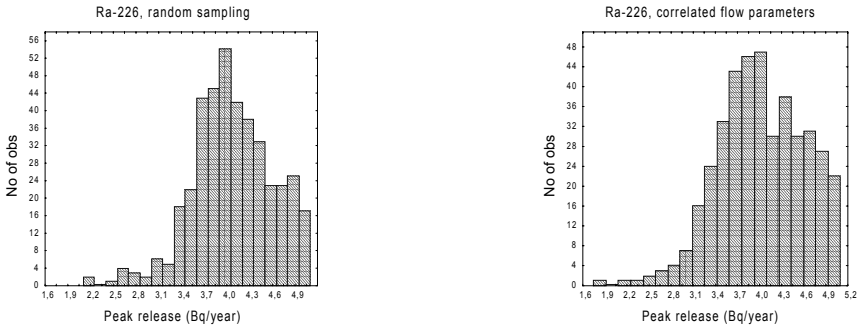

Figure 21 *The release rate histogram for Ra-226. The picture to the left corresponds to random sampling of flow parameters and that to the right to the case of correlated flow parameters.*

I-129

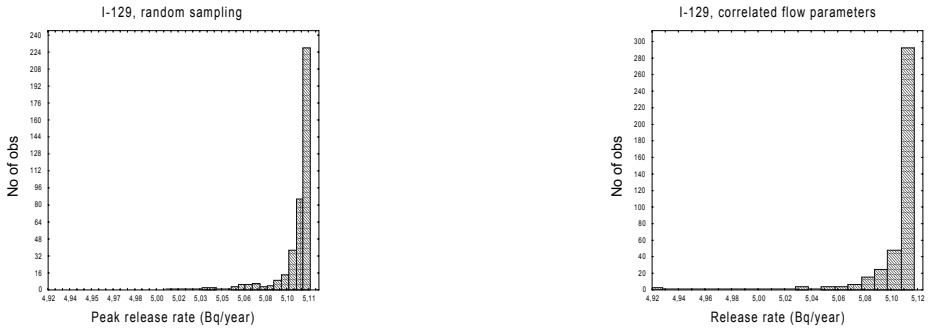

Figure 22 *The release rate histogram for I-129. The picture to the left corresponds to random sampling of flow parameters and that to the right to the case of correlated flow parameters.*

Cs-135

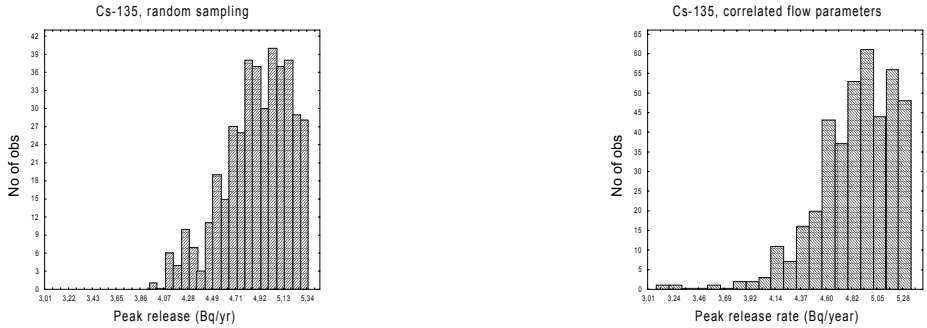

Figure 23 *The release rate histogram for Cs-135. The picture to the left corresponds to random sampling of flow parameters and that to the right to the case of correlated flow parameters.*

C-14

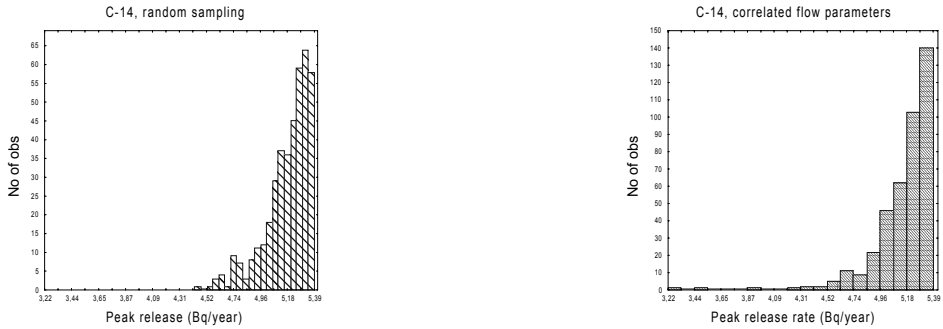

Figure 24 *The release rate histogram for C-14. The picture to the left corresponds to random sampling of flow parameters and that to the right to the case of correlated flow parameters.*

Table X *Central values for all nuclides(uncorrelated parameters).*

	C-14	Cs-135	Cl-36	I-129	Ra-226
Mean	1.62×10^{5}	9.58×10^{4}	6.56×10^{4}	1.27×10^{5}	2.28×10^{4}
Confid. (-95%)	1.57×10^{5}	9.05×10^{4}	6.56×10^{4}	1.27×10^{5}	2.01×10^{4}
Confid. (+95%)	1.67×10^{5}	1.01×10^{5}	6.56×10^{4}	1.28×10^{5}	2.54×10^{4}
Median	1.68×10^{5}	8.52×10^{4}	6.56×10^{4}	1.29×10^{5}	1.12×10^{4}
Lower Quart.	1.25×10^{5}	5.25×10^{4}	6.52×10^{4}	1.27×10^{5}	5.59×10^{3}
Upper Quart.	2.05×10^{5}	1.35×10^{5}	6.57×10^{4}	1.30×10^{5}	2.75×10^{4}
Std. Dev.	5.29×10^{4}	5.41×10^{4}	6.71×10^{1}	3.81×10^{3}	2.71×10^{4}

The differences between the shape of the output distributions resulting from the Monte Carlo calculations with and without correlations is most striking for the Cl-36 nuclide. The mean values of the peak releases are close. The correlations between the two flow parameters do not impact in a significant way on the mean peak releases. The uncertainties are lower than for the case of correlated parameters, but the reduction in the variance is almost insignificant.

5.2.2 Sensitivity Analysis

A comparison of the results of the sensitivity analysis of the case correlated flow parameters (Table IX) with the corresponding results for uncorrelated flow parameters (Table XI) reveals some very interesting features for all nuclides in question:

Cs-135, C-14: in Table IX the longitudinal dispersion D_L is very important, but in Table XI it is totally irrelevant according to the Spearmans test; there is no difference for the Pearssons test.

Cl-36: the same is true as above, but with all correlations having lower numerical values.

Ra-226: As far as the longitudinal dispersion is concerned, this is important in both cases as shown for the two tables (correlated and uncorrelated); and as far as the wetted surface area is concerned, the wetted surface area is very important for the correlated case (Table IX) but not for the uncorrelated one (Table XI).

I-129: the longitudinal dispersion is important for the correlated case as identified by the Spearman test; the other results indicate the longitudinal dispersion as being insignificant for both the correlated and uncorrelated cases.

Table XI *Results of the sensitivity analysis with uncorrelated flow parameters. The integers in bold represent the parameter rank.*

	Cs-135				**C-14**				**Ra-226**			
	Peak release				Peak release				Peak release			
Parameter	Pearsson		Spearman		Pearsson		Spearman		Pearsson		Spearman	
Spearman												
Q [m/year]	0.36	**2**	0.42	**2**	0.36	**2**	0.41	**2**	0.21	**3**	0.38	**2**
K_d [m^3/kg]	-0.19	**3**	-0.22	**3**	-0.32	**3**	-0.29	**3**	-0.69	**1**	-0.76	**1**
D_L [m^2/year]	0.17	**4**	0.08	**4**	0.12	**4**	0.06	**4**	0.27	**2**	0.31	**3**
P_{depth} [m]	0.01	**5**	0.01	**5**	-0.03	**5**	-0.04	**5**	-0.02	**5**	-0.03	**4**
a [1/m]	-0.82	**1**	-0.82	**1**	-0.76	**1**	-0.79	**1**	-0.09	**4**	-0.15	**5**

	Cl-36				**I-129**			
	Peak release				Peak release			
Parameter	Pearsson		Spearman		Pearsson		Spearman	
Q [m/year]	0.17	**4**	0.14	**3**	0.24	**3**	0.36	**3**
K_d [m^3/kg]	-0.24	**3**	-0.14	**3**	-0.24	**3**	-0.19	**4**
D_L [m^2/year]	0.08	**5**	-0.05	**4**	0.05	**4**	0.01	**5**
P_{depth} [m]	-0.26	**2**	-0.17	**2**	-0.36	**2**	-0.37	**2**
a [1/m]	-0.31	**1**	-0.31	**1**	-0.49	**1**	-0.74	**1**

6. Summary and conclusions

A number of Monte Carlo simulations of radionuclide migration in the geosphere have been done and are reported here. We have used specific and generic data. Part of the data comes from the SKI's SITE-94 exercise. The radionuclides examined were C-14, Cs-135, Cl-36, I-129 and Ra-226.

A new procedure for the transfer of hydrogeological data to a geosphere transport model of radionuclides has been tested. This procedure is more direct than the one usually used in Monte Carlo calculations of radionuclide transport. It also preserves the structure of the input data and consequently eventual correlations that exist between parameters, which is not the case with the conventional method. In the present report the only data in which such correlations were presented is that for the groundwater flow rate, i.e. the correlations are between the Darcy velocity and the longitudinal dispersion.

It has been observed that the spatial variability of these parameters is of two orders of magnitude for the Darcy flow and five orders of magnitude for the longitudinal dispersion. The parameters propagate through the geosphere model in such a way that the uncertainty of the output distribution of the peak release rates is very low. This situation is valid if no parameters other than the two mentioned above are varied in the Monte Carlo simulations.

Monte Carlo simulations in which both the flow and the transport parameters are varied show considerable spread in the distribution of the peak release rate. Hence the spread of peak releases is more influenced by the uncertainty of the transport parameters and less by the spatial variability of the flow parameters. One should, however, bear in mind that there is a synergetic effect between flow and transport phenomena, the impact of which is difficult to disentangle.

A MatLab program was written with the aim of extracting random numbers from a joint distribution. Using this program it is now possible to sample an arbitrary number of random numbers from a limited data set whilst retaining the correlation structure of the original data. With the help of this program it was possible to show that 406 samples of data taken from the hydrogeological modelling was sufficient to get the convergence of the Monte Carlo simulations of peak releases if the number of variable parameters entering the CRYSTAL model is limited to the following: Darcy velocity, longitudinal dispersion, wetted surface area, penetration depth and distribution coefficients.

The impact of the above mentioned correlations on the distribution of the peak release rates was studied. It was found that these correlations have no appreciable impact on the mean peak value of the maximum releases.

We demonstrate that the existence of correlations between parameters makes it difficult to obtain sensitivity indices that are reliable. In fact it was observed that the global methods of performing sensitivity analysis in current use break down in the presence of correlations, i.e., it is not possible to disentangle the influence of correlated parameters on the output distribution with known methods. This area requires the development of new sensitivity methods.

We have also performed some development on a new method of sensitivity analysis and got promising results for the case of correlated parameters, as shown for the simulations of Cs-135 using 4060 correlated samples. The method is based on a neural networks approach, is global, robust and computationally inexpensive. However, there is a need to develop our method further and also to design software tools to facilitate its use.

References

Andersson, K., Chemical and Physical Transport Parameters for SITE-94. SKI Report **96:2**, February 1996, Stockholm, Sweden.

Geier, J. E., Discrete-feature modelling of the Äspö site:3. Predictions of hydrogeological parameters for performance assessment (SITE-94), SKI Report **96:7**, Swedish Nuclear Power Inspectorate, Stockholm, 1996.

Helton J. C., *Risk Analysis* **14**, 483-511 (1994).

Helton J. C., Andersson D. R., Baker B. L., Bean J. E., Berglund J. W., Beyeker W., Economy K., Garner J. W., Hora S. C., Iuzzolino H. J., Knupp P., Marietta M. G., Rath J., Rechard R. P., Roache P. J., Rudeen D. K., Salari K., Schreiber J. D., Swift P. N., Tierney M. S. and Vaughn P., *Reliability Engineering and System Safety* **51**, 53-100 (1996).

Iman R. L. and Conover W. J., *Communications in Statistics*, **B11**(3), 311-334 (1982).

Montoto, M., Heath M. J., Rodriguez Rey, A., Ruiz de Argandoña, V. G., Calleja, L., and Menéndez, B., Natural analogue and microstructural studies in relation to radionuclide retardation by rock matrix diffusion in granite. Final Report. EUR 14352 EN, CEC, 1992.

Neretnieks, I., Matrix diffusion. How confident are we? In Sixth EC Natural Analogue Working Group Meeting, Proc. Int. Workshop in Santa Fe, 12-16 September 1994, (H. Von Maravic and J. Smellie, eds.), EUR 16762 EN, European Commission, Luxembourg, 1996.

Paté-Cornell M. E., *Reliability Engineering and System Safety* **54**, 95-111 (1996).

Pagis, CEC, Directorate General for Science, Research and Development, Joint Research Centre, **EUR 11775 EN** (1988).

Pereira, A., SKI TR 99:11, Spatial Variability of Geosphere Parameters and the Impact of the F-ratio on Radionuclide Release, March 1999.

Pereira, A., and Sundström, B., A Method to Induce Correlations between Distributions of Input Variables for Monte Carlo Simulations, *Radioactive Waste Management and Environmental Restoration Journal*, accepted (2000).

Prado, P., Draper, D., Saltelli, A., Pereira, A., Mendes, B., Eguillor, S., Cheal, R., and Tarantola, S., GESAMAC- Conceptual and Computational tools to tackle Long-term Risk from Nuclear Waste disposal in the Geosphere. Nuclear Science and Technology **EUR 19113**, Euratom, (1999).

Robinson, P. and Worgan, K., CRYSTAL: A model of a fractured rock geosphere for performance assessment within SKI Project-90, , SKI Technical Report **91:13**, Stockholm, Sweden (1992).

Saltelli, A., and Bolado, R., An Alternative Way to Compute Fourier Amplitude Sensitivity Test (FAST), *Computational Statistics and Data Analysis*, **26**, 445-460 (1998).

SKB 91, SKB Technical Report **92-20**, Sweden (1992).

SKI Project-90, Volumes I and II, SKI **TR 91:23**, Swedish Nuclear Power Inspectorate, Stockholm (1991).

SKI SITE-94. Deep Repository Performance Assessment Project. 1996. SKI Report **96:36**. Sweden.

Sobol', I. M., Sensitivity Analysis for Nonlinear Mathematical Models, *Mathematical Modelling & Computational Experiment*, **1**, 407- 414 [translation of Sobol (1990), Sensitivity Estimates for Nonlinear Mathematical Models, *Matematicheskoe Modeliovanie*, **2**, 112-118 (in Russian)], (1993).

SYVAC/SU, Swedish Nuclear Power Inspectorate, Stockholm, Sweden.

TVO, TVO-92 Safety analysis of spent fuel disposal, Report YJT-92-33 E, Nuclear Waste Commission of Finnish Power Companies, Helsinki, 1992.

Worgan, K., and Robinson, P., The CALIBRE Source term code: Technical Documentation for Version 2, SKI Report **95:13**, Swedish Nuclear Power Inspectorate, Stockholm, 1995b.

Appendix I. Near-field and far-field models of SYVAC/SU

The near-field model CALIBRE

CALIBRE is a 2D near-field model. The master equations of CALIBRE are given below for the sake of completeness. For details, including a discussion of the choice of boundary conditions, the reader should consult Worgan and Robinsson, (1995).

Transport in the waste-form (completely failed canister[2]):

$$\frac{\partial A_{ij}^{ca}(r,z,t)}{\partial t} = -\lambda_{ij} A_{ij}^{ca}(r,z,t) + \lambda_{IJ} A_{IJ}^{ca}(r,z,t) + D_e^B \left[\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial c_{ij}^{ca}(r,z,t)}{\partial r} \right) \right] + D_e^{ca} \frac{\partial^2 c_{ij}^{ca}(r,z,t)}{\partial z^2}$$

Transport in the bentonite:

$$\frac{\partial A_{ij}^{B}(r,z,t)}{\partial t} = -\lambda_{ij} A_{ij}^{B}(r,z,t) + \lambda_{IJ} A_{IJ}^{B}(r,z,t) + D_e^B \left[\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial c_{ij}^{B}(r,z,t)}{\partial r} \right) + \frac{\partial^2 c_{ij}^{B}(r,z,t)}{\partial z^2} \right]$$

Transport in the near-field rock:

$$\frac{\partial A_{ij}^{ro}(r,z,t)}{\partial t} = -\lambda_{ij} A_{ij}^{ro}(r,z,t) + \lambda_{IJ} A_{IJ}^{ro}(r,z,t) + D_e^{ro} \left[\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial c_{ij}^{ro}(r,z,t)}{\partial r} \right) + \frac{\partial^2 c_{ij}^{ro}(r,z,t)}{\partial z^2} \right]$$

Transport in the fracture:

$$\frac{\partial A_{ij}^{fr}(r,z,t)}{\partial t} = -\lambda_{ij} A_{ij}^{fr}(t) + \lambda_{IJ} A_{IJ}^{fr}(t) + D^w \left[\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial c_{ij}^{fr}}{\partial r} + \frac{\partial^2 c_{ij}^{fr}}{\partial z} + \frac{1}{r^2} \frac{\partial^2 c_{ij}^{fr}}{\partial \phi^2} \right) \right] + \vartheta_r \frac{\partial c_{ij}^{fr}}{\partial r} + \vartheta_\phi \frac{1}{r} \frac{\partial c_{ij}^{fr}}{\partial \phi}$$

[2] CALIBRE models also pinhole release

The far-field model CRYSTAL

The transport in the far-field is modelled by CRYSTAL. The equations given below are presented for the sake of completeness and the reader should consult Robinson and Worgan (1992) for details, including the treatment of boundary conditions. The transport is described by the system of partial differential equations:

$$R_n \frac{\partial C_n}{\partial t} = -u \frac{\partial C_n}{\partial x} + D \frac{\partial^2 C_n}{\partial x^2} + a \frac{\theta^m D^m}{\theta} \frac{\partial C_n^m}{\partial \omega}\bigg|_{\omega=0} - R_n \lambda_n C_n + R_{n-1} \lambda_{n-1} C_{n-1}$$

$$R_n^m \frac{\partial C_n^m}{\partial t} = D^m \frac{\partial^2 C_n^m}{\partial \omega^2} - R_n^m \lambda_n C_n^m + R_{n-1}^m \lambda_{n-1} C_{n-1}^m$$

where R_n is the retardation attributable to sorption on the fracture walls and R_n^m the retardation from the matrix; these retardation coefficients are given by $R_n = 1 + \rho^m K_{d,n}(1-\theta^m)a\delta/\theta$ and $R_n^m = 1 + K_{d,n}\rho^m(1-\theta^m)/\theta^m$ respectively. The subscript n denotes the n^{th} radionuclide in the chain. $C_n(x.t)$ (moles/m^3) is the concentration of radionuclides in the fracture water; t (years) is the time; u (m/year) is the velocity of water in the fracture; x (m) is the distance along the pathway; D (m^2/year) is the longitudinal dispersion; D^m (m^2/year) is the matrix diffusivity; θ^m is the matrix porosity and θ the rock mass porosity; a (m^{-1}) is the specific flow-wetted area per volume of the rock mass, δ (m) is the "depth" of surface sorption; $K_{d.n}$ (m^3/kg) is the distribution coefficient; $C_n^m(x,\omega,t)$ (moles/m^3) is the concentration of radionuclides in the rock matrix pore water; ω (m) is the distance perpendicular to the fracture; δ (m) is the maximum penetration depth in the ω direction and λ (year)$^{-1}$ is the radioactive decay constant.

Appendix II. Exploratory calculations of Cs-135

In this appendix we study seven cases where the two flow parameters Darcy flow and hydrodynamic dispersion are always given by the bivariate *pdf* and the other three parameters are varied one at a time or using different combinations. These cases are: a) Monte Carlo calculations with three varying parameters where the third variable parameter is the wetted specific surface; b) Monte Carlo calculations with three variable parameters, but now with the distribution coefficient as the third variable parameter; c) Monte Carlo calculations with three variable parameters where the third one is the penetration depth; d) Monte Carlo calculations where the third and fourth parameters are the distribution coefficient and the specific wetted surface; area e) Monte Carlo calculations where the third and fourth parameters are the penetration depth and the specific wetted surface and f) Monte Carlo calculations where five parameters are varied simultaneously, i.e., the Darcy velocity, the hydrodynamic dispersion, the specific wetted surface area, the distribution coefficient and the penetration depth.

The input data for these different cases are given by Table I of Section 4, with the exception of the parameters which are varied here, their values being given in Table II-1 of this appendix. These parameters are the specific wetted surface area a, the penetration depth P_{depth} and the distribution coefficient K_d. The *pdfs* chosen for these parameters are uniform and the ranges (minimum and maximum values) are taken from the SITE-94 report. The specific wetted surface is bounded by the minimum and maximum values given in Table 15.2.3, Vol. II of the SITE-94 report (1996). The penetration depth (range of matrix diffusion) observed in natural analogues (Montoto et al., 1992 and Neretnieks, 1996) in granite ranges from 2-10 cm. These figures are for highly adsorbing nuclides and therefore a upper bound of 25 cm is used in SITE-94. The range given in Table II-1 is therefore 2-25 cm. The distribution coefficient of caesium ranges from 0.1-0.5 m^3/kg (Andersson, 1996).

The results of these calculations are shown as histograms for the peak releases in Figure II-1 of this Appendix and the uncertainties are summarised in Table II-2. This table is organised according to increasing uncertainty (see standard deviation) from the left column to the right one, i.e. from Case I to Case VII (see also Figure II-2). Both the mean value and the median show a tendency to decrease as the uncertainty becomes larger.

Figure II-2 shows the propagation of the uncertainties with different numbers of varying parameters according to the cases shown in Table II-1. Figure II-3 shows uncertainties in the peak releases for Case V together with those of the input parameters for ^{135}Cs. It becomes clear that the unusually small uncertainties apparent in Section 4 where only two flow parameters were allowed to vary increases now by more than one order of magnitude. These uncertainties are nevertheless considerably smaller than those obtained in the past in several probabilistic calculations or assessment exercises, such as Pagis (1998), Project-90 (1991), SKB91 (1992) and GESAMAC (1999).

Table II-1 Values of far field parameters *for the constant parameters.*

Case	Parameter	PDF	Value	Unit
1	Penetration depth. P_{depth}	Uniform	2.0E-02. 2.5E-01	m
2	Distribution coefficient. K_d	Uniform	1.0E-01. 5.0E-01	m^3/kg
3	Specific wet. area. a	Uniform	6.3E-03. 7.86E-01	1/m
4	Distribution coefficient. K_d	Uniform	1.0E-01. 5.0E-01	m^3/kg
	Specific wet. area. a	Uniform	6.3E-03. 7.86E-01	1/m
5	Distribution coefficient. K_d	Uniform	1.0E-01. 5.0E-01	m^3/kg
	Penetration depth. P_{depth}	Uniform	2.0E-02. 2.5E-01	m
6	Specific wet. area a	Uniform	6.3E-03. 7.86E-01	1/m
	Penetration depth. P_{depth}	Uniform	2.0E-02. 2.5E-01	m
7	Specific wet. area. a	Uniform	6.3E-03. 7.86E-01	1/m
	Distribution coefficient. K_d	Uniform	1.0E-01. 5.0E-01	m^3/kg
	Penetration depth. P_{depth}	Uniform	2.0E-02. 2.5E-01	m

Table II-2 *Central values for caesium variation calculations organised in increasing order of uncertainty (standard deviation) .*

	Cs-135						
	Case I	**Case V**	**Case II**	**Case III**	**Case VI**	**Case IV**	**Case VII**
Variable parameters	P_{depth}	$K_d + P_{depth}$	K_d	a	$a + P_{depth}$	$K_d + a$	$K_d + a + P_{depth}$
Mean	2.17×10^5	2.09×10^5	2.10×10^5	1.16×10^5	1.17×10^5	8.99×10^4	9.41×10^4
Confid. (-95%)	2.16×10^5	2.08×10^5	2.09×10^5	1.11×10^5	1.12×10^5	8.47×10^4	8.85×10^4
Confid. (+95%)	2.18×10^5	2.10×10^5	2.11×10^5	1.21×10^5	1.22×10^5	9.52×10^4	9.96×10^4
Median	2.19×10^5	2.11×10^5	2.13×10^5	1.09×10^5	1.16×10^5	8.24×10^4	8.26×10^4
Minimum	1.65×10^5	1.48×10^5	1.27×10^5	2.42×10^4	4.45×10^3	6.41×10^4	1.23×10^3
Maximum	2.29×10^5	2.29×10^5	2.29×10^5	2.29×10^5	2.29×10^5	2.28×10^5	2.28×10^5
Lower Quart.	2.14×10^5	2.04×10^5	2.06×10^5	7.67×10^4	7.62×10^4	4.57×10^4	4.56×10^4
Upper Quart.	2.22×10^5	2.17×10^5	2.18×10^5	1.50×10^5	1.55×10^5	1.28×10^5	1.36×10^5
Std. Dev.	7.40×10^3	1.08×10^4	1.12×10^4	4.94×10^4	5.17×10^4	5.40×10^4	5.69×10^4

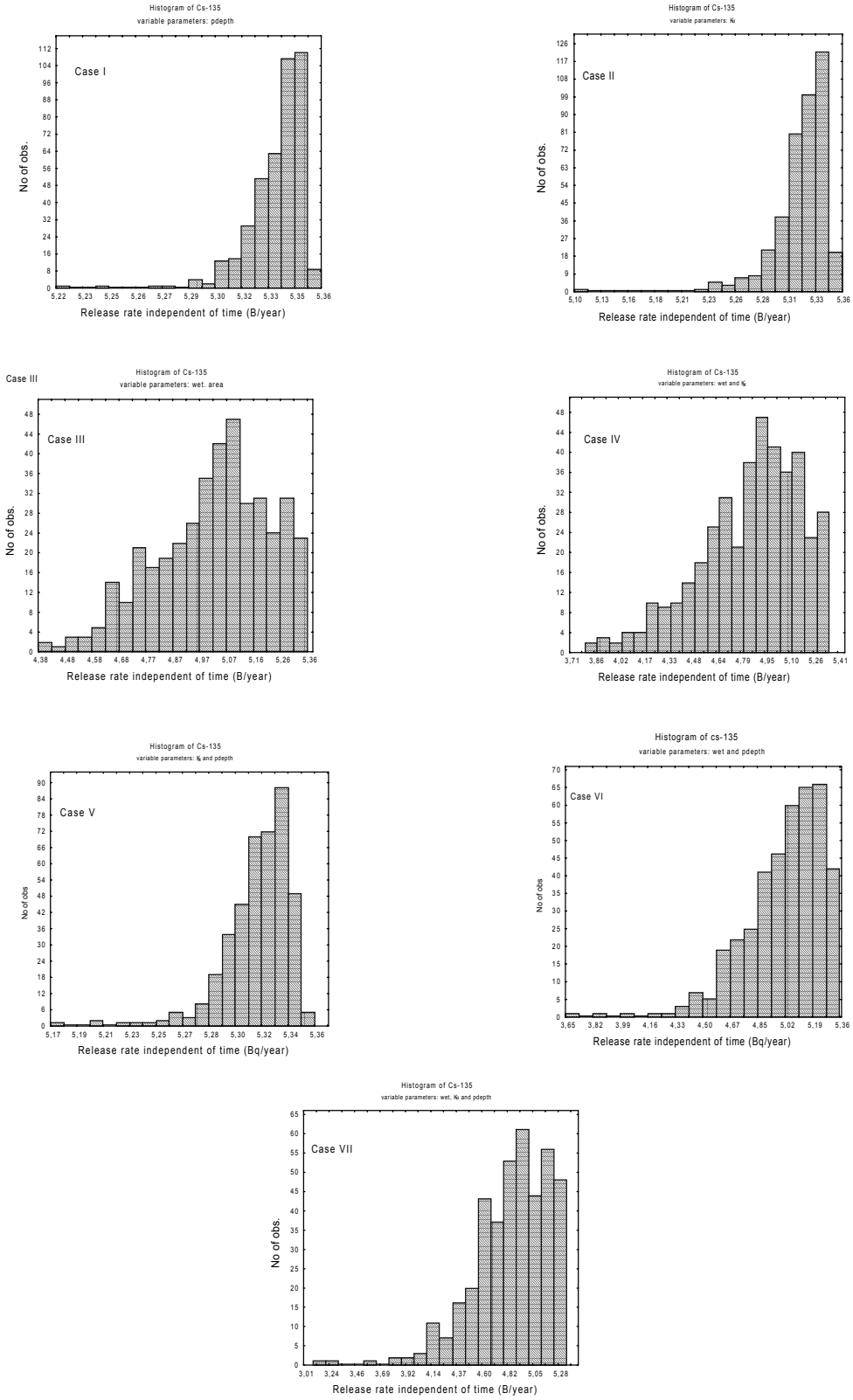

Figure II-1 *The release rate histogram for Cases I to Case V shows an increase in the variance with an increasing number of variables.*

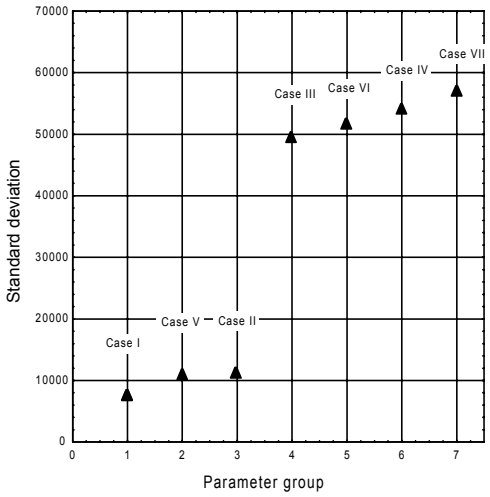

Figure II-2 *The increase in the standard deviation of the peak release of the ^{135}Cs peak release distribution with the increasing number of variables.*

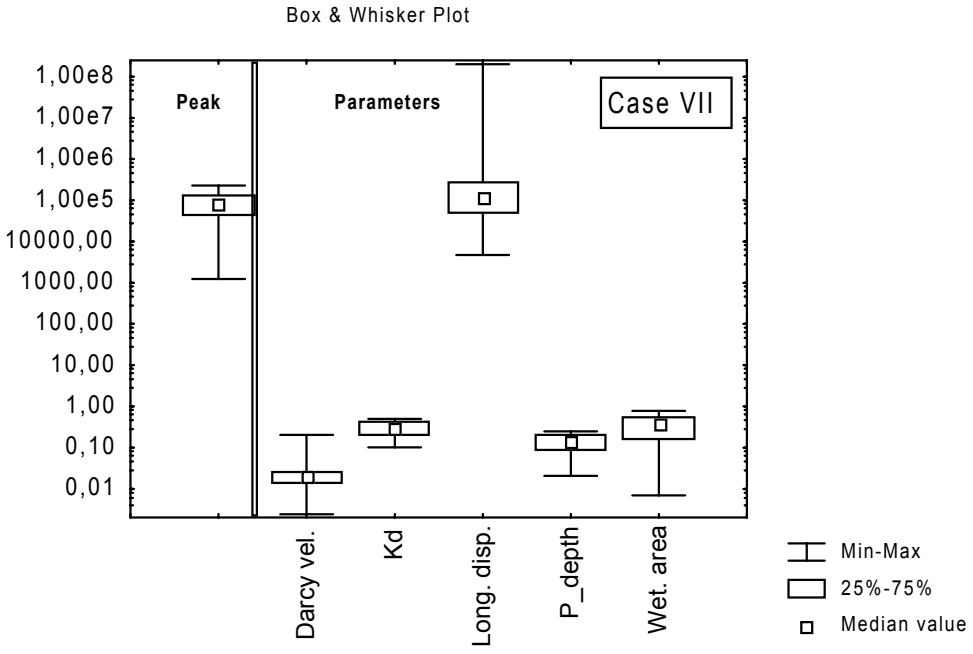

Figure II-3 *The median of the peak release of the ^{135}Cs and the quartiles is shown at left in the figure. At right, the same quantities are shown for each of the varying parameters.*

Appendix III. A program to sample correlated random numbers

The aim of the following MatLab code is to sample random numbers from a joint probability distribution. This joint distribution is given by a bivariate histogram, which may be obtained from experimental results or from simulations.

The code is a useful tool that retains the correlation structure of the input data while generating an arbitrarily high number of samples for use in Monte Carlo simulations whenever many realisations required to obtain sufficiently good statistics. The code has been used in this work wherever the number of parameters to be sampled needed more than 406 realisations, which was the number of available "data points" from hydrogeological calculations.

```
clear all; clear; close all;
'Bellow, the limits to the Statistica-plots are given'

xmin=input('Give the minimum value of the vector x: ')
xmax=input('Give the maximum value of the vector x: ')
ymin=input('Give the minimum value of the vector y: ')
ymax=input('Give the maximum value of the vector y: ')
f=input('Give factor f (integer): ')

% Substitute the matrix shown bellow by your frequency-
% matrix, where row 1 to row(max) corresponds to x(1) to
% x(max) and colon  1 to colon(max) corresponds to y(1) to
% y(max)

M=[
2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0
0  1  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
1  2  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  2  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  6  6  4  2  4  0  1  0  0  0  0  0  0  0  0  0  0  0  0
0  1  6  5  6  2  3  1  3  0  0  0  0  0  0  0  0  0  0  0
0  0  6  12 13 7  6  0  2  0  0  0  0  0  0  0  0  0  0  0
0  0  2  10 16 16 13 2  2  1  0  0  0  0  0  0  0  0  0  0
0  0  0  7  8  13 17 15 5  1  0  0  0  0  0  0  0  0  0  0
0  0  0  3  7  12 12 13 18 4  1  0  0  0  0  0  0  0  0  0
0  0  0  0  4  5  10 13 11 3  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  2  3  14 7  5  1  2  0  0  0  0  0  0  0  0
0  0  0  0  0  1  1  1  3  4  1  0  0  0  0  0  0  0  0  0
0  0  0  0  0  1  0  2  1  1  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
```

```
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  2
];

intervall=length(M(1,:));
r=intervall;
k=intervall;

% transpose the matrix to make the indexing easier
%-----------------------------------------------

M=M';
m1=sum(M);
count=sum(m1);

% Compute the bins of x and y
%----------------------------------------

dx=abs((xmax-xmin)/(intervall-2));
x=[xmin-dx];
dy=abs((ymax-ymin)/(intervall-2));
y=[ymin-dy];

for i=1:(intervall)
   x=[x x(i)+dx];
   y=[y y(i)+dy];
end

bins_x=x;
bins_y=y;

% Compute the new values in the vectors xny and yny
%--------------------------------------------------

xny=[];
yny=[];
for radnr=1:20
   for kolnr=1:20
      for i=1:(f*M(radnr,kolnr))
         if M(radnr,kolnr)==0
            break
            else
         xny=[xny (bins_x(kolnr)+rand*(bins_x(kolnr+1)-
bins_x(kolnr))) ];
        yny=[yny (bins_y(radnr)+rand*(bins_y(radnr+1)-
bins_y(radnr))) ];
      end
      end
   end
```

```
end


% Print a check on the number of elements (before and after)
%-------------------------------------------------------------

erhållet_antal_element=length(xny(1,:))
ursprungligt_antal_element=count
'f * ursprungligt_antal_element =', f*count

% Create the matrix Mny with the new frequency-distribution
%-------------------------------------------------------------

Mny=ones(r,k);
for i = 1:k
   for j=1:r
      xi=find( (xny <= bins_x(i+1)) & (xny > bins_x(i)));
      yi=find( (yny <= bins_y(j+1)) & (yny > bins_y(j)));
      Mny(j,i)=length(intersect(xi,yi));
   end
end
side=1:intervall;
Mny=Mny';
colormap('summer')
bar3(side,Mny)
xlabel('Vector y')
ylabel('Vector x')
zlabel('Frequency')
```

Monte Carlo calculations were performed for the radionuclide Cs-135 using 4060 pairs of data points (representing the Darcy velocity and longitudinal dispersion). This data is given by a joint distribution, the histogram of which is shown in Figure III-1 and is essentially the same as the one we started with. The central values for the peak release rates of Cs-135 are given by Table III-1. From this table one concludes that the mean values for the peak release using 406 samples is sufficiently near to the one obtained in Monte Carlo calculations with the number of samples one order of magnitude higher. Hence the hydrogeological data is both complete and sufficient (see Chapter 3.3) for the purpose of performing Monte Carlo transport calculations whenever only five variable parameters enter into them and it is not necessary to recalculate the cases studied in Chapters 4 and 5.

Table III-1 *Central values for Cs-135 peak releases - 4060 correlated samples.*

Nuclide	**Cs-135**
Mean	9.01×10^4
Confid. (-95%)	8.85×10^4
Confid. (+95%)	9.17×10^4
Median	7.91×10^4
Lower Quart.	4.84×10^4
Upper Quart.	1.25×10^5
Std. Dev.	5.34×10^4

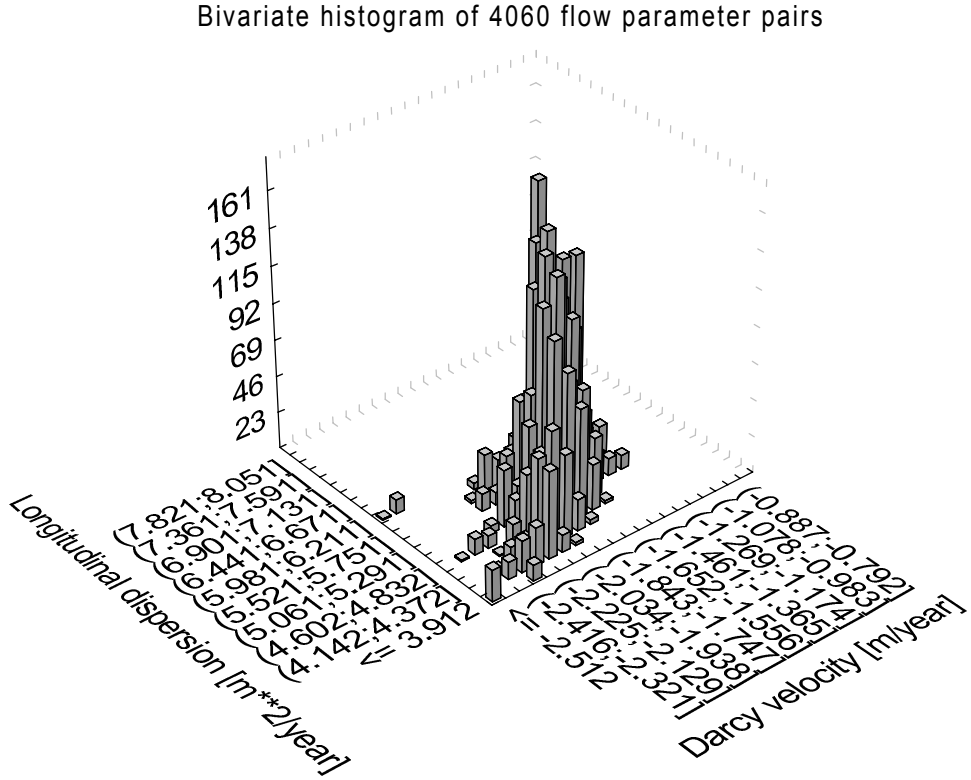

Figure III-1 *The generated joint distribution of the Darcy velocity and the longitudinal dispersion represented by a histogram of 4060 samples (one order of magnitude higher than the number of data pairs obtained from the hydrogeological simulations).*

Appendix IV. On the sensitivity analysis of models with correlated variables

This appendix summarises some preliminary results for a new method of global sensitivity analysis that is robust, computationally fast and expected to cope with correlated parameters. The method is based on a neural network approach.

Sensitivity results were obtained for flow and transport parameters governing the migration of Cs-135 in the geosphere. The sensitivity values and the ranking of the variable parameters in order of the importance their impact has on the output distribution is shown in Tables IV-1 and IV-2 below. The number of Monte Carlo simulations used was 4060. The stability of the approach is shown by using four different types of neural network: 10-5-1, 15-5-1, 20-5-1 and 10-10-1. In the 10-5-1 configuration, the first number (10) gives the number of input nodes of the neural network, 5 gives the number of hidden nodes and 1 the number of output nodes. It is concluded from this preliminary work that an important condition for reliable ranking is the success of the network in predicting the release rates. A low error is *necessary* to be certain that the correct answer is obtained.

Excluding the results from the neural networks that give poor neural networks we find a very good agreement between the results obtained from the different configurations. Even for the "bad" ANNs (artificial neural networks) we can see an agreement if the increments (which are not defined here) are equal or higher than 9%.

Table IV-1 *Sensitivity analysis results for Cs-135 with two correlated parameters.*

Configuration	Parameter	1 %	4 %	9 %	16 %	25 %
10-5-1	Darcy	0.0024	0.0112	0.0324	0.0723	0.1312
	K_d	0.0020	0.0020	0.0022	0.0026	0.0032
	Disp	0.1071	0.8829	2.0160	2.8890	2.9703
	Pdepth	0.0019	0.0018	0.0017	0.0016	0.0014
	Wetarea	0.0021	0.0033	0.0075	0.0164	0.0306
15-5-1	Darcy	0.0006	0.0096	0.0313	0.0718	0.1305
	K_d	0.0001	0.0001	0.0004	0.0007	0.0013
	Disp	0.1039	0.9872	2.3606	3.0065	2.6556
	Pdepth	0.0001	0.0002	0.0003	0.0004	0.0005
	Wetarea	0.0002	0.0012	0.0053	0.0139	0.0277
20-5-1	Darcy	0.0005	0.0096	0.0310	0.0713	0.1308
	K_d	0.0001	0.0001	0.0004	0.0007	0.0013
	Disp	0.1007	0.8357	1.9793	3.6191	3.7686
	Pdepth	0.0001	0.0001	0.0002	0.0003	0.0004
	Wetarea	0.0001	0.0012	0.0053	0.0139	0.0277
10-10-1	Darcy	0.0006	0.0097	0.0313	0.0718	0.1302
	K_d	0.0002	0.0002	0.0004	0.0008	0.0014
	Disp	0.1143	1.2504	3.3751	7.1720	8.1421
	Pdepth	0.0001	0.0002	0.0002	0.0003	0.0004
	Wetarea	0.0002	0.0014	0.0054	0.0141	0.0279

Table IV-2 shows that the parameters should be ranked in the following way:

$D_L > q > a > K_d > P_{depth}$
where D_L is the longitudinal dispersion, q the Darcy velocity, a the wetted surface area, K_d is the distribution coefficient and P_{depth} is the penetration depth of radionuclides in the matrix.

Table IV-2 *Parameter ranking where 1 is the most important parameter (correlated parameters).*

Configuration	Error	Parameter	1 %	4 %	9 %	16 %	25 %
10-5-1	0.07	Darcy	2	2	2	2	2
		K_d	4	4	4	4	4
		Disp	1	1	1	1	1
		Pdepth	5	5	5	5	5
		Wetarea	3	3	3	3	3
15-5-1	0.2	Darcy	2	2	2	2	2
		K_d	5	5	4	4	4
		Disp	1	1	1	1	1
		Pdepth	4	4	5	5	5
		Wetarea	3	3	3	3	3
20-5-1	0.1	Darcy	2	2	2	2	2
		K_d	5	5	4	4	4
		Disp	1	1	1	1	1
		Pdepth	4	4	5	5	5
		Wetarea	3	3	3	3	3
10-10-1	0.06	Darcy	2	2	2	2	2
		K_d	4	4	4	4	4
		Disp	1	1	1	1	1
		Pdepth	5	5	5	5	5
		Wetarea	3	3	3	3	3