EUROPEAN COMMISSION
Community research

PAMINA
Performance Assessment Methodologies in Application to Guide the Development of the Safety Case

(Contract Number: FP6-036404)

SENSITIVITY/UNCERTAINTY ANALYSES
APPLICATION TO A REPOSITORY IN GRANITE
MILESTONE (N°: M2.1.D.8)

Author(s):

José L. Cormenzana López
Enresa

Ricardo Bolado Lavin
JRC Petten

Date of issue of this report : **04/06/2009**

Revision: 1

Start date of project : **01/10/2006**

Duration : 36 Months

Project co-funded by the European Commission under the Euratom Research and Training Programme on Nuclear Energy within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
RE	Restricted to a group specified by the partners of the [PAMINA] project	
CO	Confidential, only for partners of the [PAMINA] project	

Foreword

The work presented in this report was developed within the Integrated Project PAMINA: **P**erformance **A**ssessment **M**ethodologies **IN** **A**pplication to Guide the Development of the Safety Case. This project is part of the Sixth Framework Programme of the European Commission. It brings together 25 organisations from ten European countries and one EC Joint Research Centre in order to improve and harmonise methodologies and tools for demonstrating the safety of deep geological disposal of long-lived radioactive waste for different waste types, repository designs and geological environments. The results will be of interest to national waste management organisations, regulators and lay stakeholders.

The work is organised in four Research and Technology Development Components (RTDCs) and one additional component dealing with knowledge management and dissemination of knowledge:

- In RTDC 1 the aim is to evaluate the state of the art of methodologies and approaches needed for assessing the safety of deep geological disposal, on the basis of comprehensive review of international practice. This work includes the identification of any deficiencies in methods and tools.
- In RTDC 2 the aim is to establish a framework and methodology for the treatment of uncertainty during PA and safety case development. Guidance on, and examples of, good practice will be provided on the communication and treatment of different types of uncertainty, spatial variability, the development of probabilistic safety assessment tools, and techniques for sensitivity and uncertainty analysis.
- In RTDC 3 the aim is to develop methodologies and tools for integrated PA for various geological disposal concepts. This work includes the development of PA scenarios, of the PA approach to gas migration processes, of the PA approach to radionuclide source term modelling, and of safety and performance indicators.
- In RTDC 4 the aim is to conduct several benchmark exercises on specific processes, in which quantitative comparisons are made between approaches that rely on simplifying assumptions and models, and those that rely on complex models that take into account a more complete process conceptualization in space and time.

The work presented in this report was performed in the scope of RTDC 2.

All PAMINA reports can be downloaded from http://www.ip-pamina.eu.

Table of contents

1 Introduction and background

This document presents the work performed by ENRESA and JRC within topic 4 “Testing of sensitivity analysis methods – the Spanish programme” of Task 2.1.D “Techniques for Sensitivity and Uncertainty Analysis” of PAMINA project.

ENRESA and JRC have collaborated to test different sensitivity analysis methods on the Performance Assessment (PA) model for repository in granite. JRC has developed sensitivity analysis tools as MATLAB programmes and ENRESA has applied them to the PA model used in the probabilistic evaluation of the Spanish disposal concept in granite.

The sensitivity analysis done by ENRESA and JRC has been an iterative process. Initial results were discussed by both partners, arising new ideas and topics that were explored in new analyses. After several iterations the sensitivity analysis reached the final form that is presented in this document.

This document provides information on:

- the disposal system under study (chapter 2),
- the models and parameters used in the PA calculations (chapter 3),
- the sensitivity analyses performed and the results obtained (chapter 4), and
- the main findings of the sensitivity analysis (chapter 5).

The review of sensitivity analysis methods done by JRC within Task 2.1.D of PAMINA project /1/ provides a useful description of the sensitivity analysis methods used in this document.

2 Description of the repository in granite

ENRESA reference concept in granite is based on the disposal of spent fuel in carbon steel canisters in 500m long horizontal drifts excavated by means of tunnel boring machines. Canisters are surrounded by bentonite. Access is accomplished by means of "main drifts" which run perpendicular to the disposal drifts. The main drifts meet at a central area, which includes the required underground infrastructure. Communication between the surface and the central underground area is accomplished by means of 3 access shafts and a ramp.

The carbon steel canister measures 4.54 m in length and 0.90 m in diameter, has a thickness of 10cm, and contains 4 PWR or 12 BWR fuel elements in a subcritical configuration. After being unloaded from the reactor, the fuel elements are temporarily stored during 50 years to allow the thermal power to decay to a level at which they may be disposed of, with a thermal power of 1,200 W per canister. A total of 3,600 canisters will be required for the total inventory of spent fuel estimated for the Spanish nuclear programme.

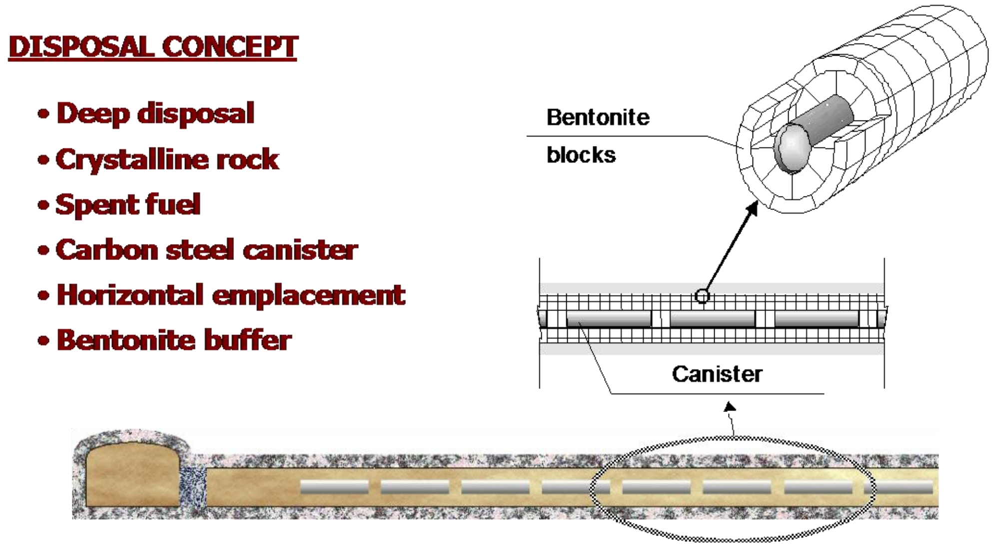


Figure 2.1 - Longitudinal section of a disposal drift

Canisters are emplaced in cylindrical disposal cells, built with blocks of pre-compacted bentonite. The disposal drifts of 500 m in length and 2.4 m in diameter (see Figure 2.1) are located at a depth of 500 m in the host formation. The separation between canisters is determined mainly by thermal constraints. Separations of 1.4 m between canisters and 35 m between disposal drifts have been established, in order not to exceed a temperature of 100ºC in the bentonite. Actual separation is a function of the properties of the host rock, and thermal calculations have been made for a reference generic site.

Once a disposal drift is filled with canisters and bentonite blocks, it is sealed with a 6 m long seal made of bentonite blocks and closed with a concrete plug at its entry. After completion of all the disposal drifts, the main drifts, ramp, shafts and other remaining rock cavities will be backfilled with a mixture of bentonite and natural sand or an appropriate crushed material. The backfilling material will consist of 10 % bentonite (increasing up to 20 % at the top of the drifts) and suitably graded sand.

3 System evolution and modelling

Due to high suction pressures in bentonite, it is expected that bentonite will be fully saturated a few decades after repository closure. When water contacts the carbon steel canister a process of anaerobic generalized corrosion starts. In the calculations performed canisters have a minimum duration of 20,000 years and fail sequentially up to 100,000 years approximately. In addition, a few canisters (10 as a maximum) are assumed to have an undetected defect and fail around one hundred years after disposal.

After canister failure, and since no credit is given to the cladding as a barrier, there is an instantaneous release of some volatile radionuclides (such as I129, Cl36 and Cs135) and the gradual release of the radionuclides in the UO_2 matrix starts.

The radionuclides released from the waste dissolve or precipitate in the water in the canister cavity, depending on their solubility limits. Dissolved radionuclides are transported by diffusion through the bentonite buffer and pass to the groundwater flowing around the disposal drifts. Transport calculations in the near field are done using a 1D axisymmetric model of the bentonite and imposing a “mixing tank” boundary condition in the bentonite outer surface, where radionuclides leave the bentonite and pass to the groundwater flowing in the granite.

In the transport calculations the granite is modelled as a single one-dimensional planar fracture (or stream tube) with a distribution of travel times provided by the hydro-geological model. Longitudinal dispersion and matrix diffusion into the adjacent rock are modelled, including sorption onto the granite. Neither solubility limits nor sorption on fracture surfaces or infill are considered.

The radionuclides released from the granite formation are discharged to a river with a flow rate of 10^6 m^3/year. Radionuclides concentrations in river water are calculated dividing the release rate form the geosphere by the river flow rate. River water is used by the critical group to produce most of the aliments they consume.

3.1 Variables used in the calculations.

In the repository model all the uncertain parameters are represented by probability distribution functions (pdf´s), with the exception of Biosphere parameters, which are constants.

The model uses 135 random input parameters, each one being represented by an independent pdf. Many of the parameters are element-specific (such as the distribution coefficients in the bentonite), and a given radionuclide can be affected by 22 parameters at most. Random input parameters are presented in Table 3.1, classified on the basis of the component of the disposal systems affected by the parameter. Element-specific parameters are shown in grey cells In Table 3.1.

Table 3.1 – List of parameters of the model potentially affecting the chemical element X.

Waste(3)	
R1000	UO_2 matrix alteration rate after 1000 years of cooling
METAL LIFETIME	Duration of the structural components of the fuel elements
GAPX	Fraction of the inventory of element X released immediately when the canister fails – Instant Release Fraction (IRF)
Canisters (3)	
FAIL	Number of canisters with initial defects that fail in a century approximately
MEAN	Mean duration of the canisters due to generalised corrosion
SLOPE	Parameter for the Weibull distribution of canister failures
Near field (6)	
VC1	Volume of water in the canister cavity
POR ANIONS BUFFER	Fraction of bentonite porosity accessible to anions
WATER NF	Groundwater flow in the granite of the near field
SOLX	Solubility of element X in the water of the canister cavity
KDBEX	Distribution coefficient bentonite-water for element X
DIFFX	Diffusion coefficient in bentonite porewater (Dp) for element X
Water flow (4)	
WATER TT0	Water travel time for a kinematic porosity of 10^{-4}
KINEM POR	Kinematic porosity of the granite
PECLET	Peclet number (longitudinal dispersion)
PATH LENGTH	Length of the geosphere pathway, from repository to biosphere
Granite matrix (6)	
FWA	Flow wetted area
THICKM	Thickness of the granite matrix
PORMAT	Porosity of the granite matrix
POR ANIONS GRANITE	Fraction of granite matrix porosity accessible to anions
MATCOR	Correction factor for applying Kd values measured in batch to intact granite
KDRX	Distribution coefficient granite-water for element X (in batch)

4 Results

Figure 4.1 shows the mean doses obtained in a stochastic calculation with 5000 runs for the Spanish disposal concept in granite. Both total doses and doses per radionuclide are displayed. The doses due to all the members of a decay chain are added and the sum is included in Figure 4.1.

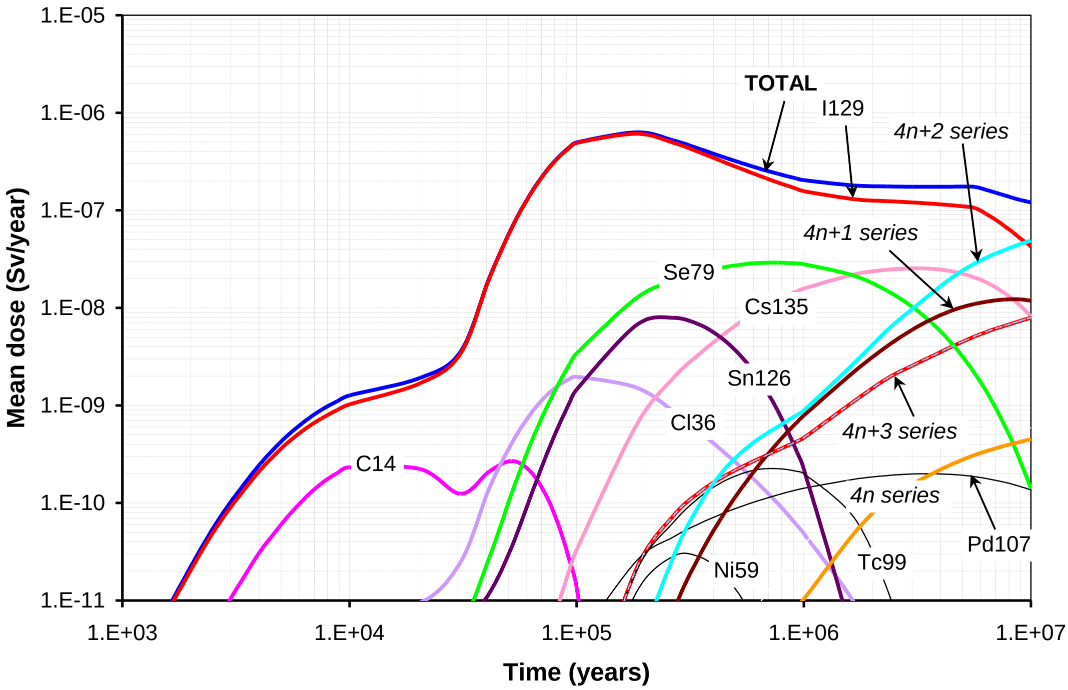


Figure 4.1 – Mean doses (total and per radionuclide) in a calculations with 5000 runs.

Figure 4.1 shows all the radionuclides that produce mean doses greater than 10^{-11} Sv/yr in the calculation (10^{-11} Sv is roughly the dose received in 0.1 seconds due to natural sources).

The sensitivity analyses done in the next sections are limited to the doses caused by the repository system, and only the radionuclides (plus total and series) presented in Figure 4.1 are considered. Nevertheless, the methodology presented in the next sections can be used to study other results of the model, such as the radionuclide fluxes leaving the near field of the repository or the time of occurrence of the peak dose.

Most analysis have been done for a simulation of 25,000 runs with Simple Random Sampling (SRS) Whenever a different simulation is used, it is stated in the text.

There are some runs with peak doses equal to zero (< $3 \cdot 10^{-31}$ Sv/yr in the calculations). The number of runs giving a dose equal to zero for each radionuclide is presented in Table 4.1.

Table 4.1 – Number of runs with zero dose in the 25,000 runs simulation.

	Number of runs With zero dose
TOTAL	0
I129	0
Cl36	0
Cs135	2180
C14	1200
Ni59	12331
Se79	0
Pd107	91
Sn126	1120
Tc99	3475
4n series	4406
4n+1 series	6779
4n+2 series	4284
4n+3 series	4355

4.1 Correlation and regression analysis for the peak doses

Due to the relevance of the peak doses in the Safety Assessment, a comprehensive correlation and regression analysis has been done.

The correlation coefficients of all the random input parameters with the total peak dose and peak dose due to a single radionuclide or a radioactive series have been calculated. The correlation coefficients have been calculated both for values (CCs) and for ranks (Rank Correlation Coefficients RCCs).

Since all the random input parameters are independent (orthogonal) the regression coefficients are equal to the equivalent correlation coefficients. The Standardized Regression Coefficients (SRCs) are equal to the Correlation Coefficients (CCs) and the Standardized Rank Regression Coefficients (SRRCs) are equal to the Rank Correlation Coefficients (RCCs).

The coefficients of determination (R^2) of the multiple regressions are calculated also. Since the random parameters are independent, the coefficient R^2 can be calculated simply as the sum of the squares of the correlation coefficients: $R^2=CC_1^2+CC_2^2+$……..

4.1.1 Initial results

4.1.1.1 Effect of the number of runs

Figure 4.2 to Figure 4.5 present the CCs and RCCs of the 135 random input parameters with the peak dose due to I129 in calculations with 200, 1000, 5000 and 25000 runs, using SRS in all the cases.

Figure 4.2 to Figure 4.5 show that the “noise” created by the non-relevant parameters decreases with the number of runs, as expected. The random input parameters with correlation coefficients greater than 0.2 (in absolute value) are identified with just 1000 runs, and the random parameters with correlation coefficients greater than 0.1 (in absolute value) are well identified with 5000 runs.

These results have been obtained for I129, but are applicable to other radionuclides also. It can be concluded that 5000 runs are enough for the analysed system and allow identifying parameters with correlation and regression coefficients greater than 0.1 (in absolute value).

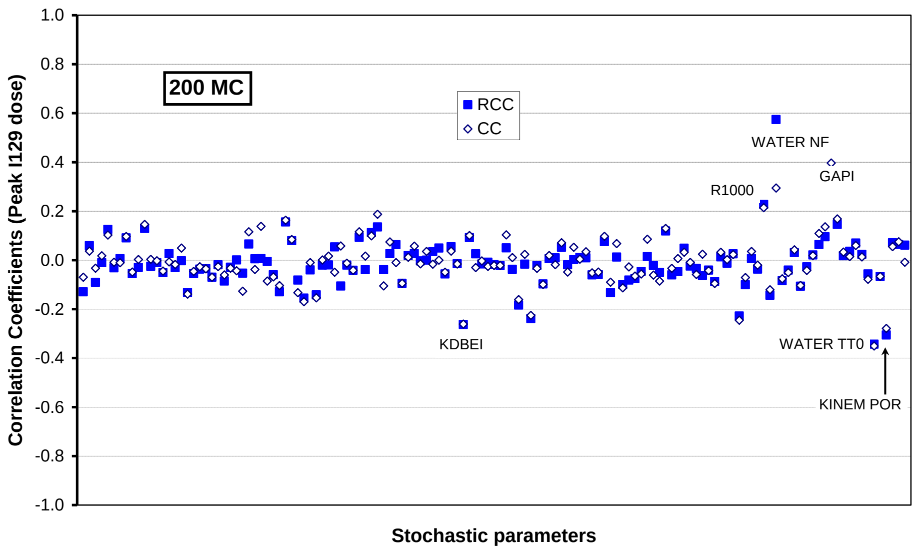


Figure 4.2 – CCs and RCCs for the peak dose due to I129 in a calculation with 200 runs.

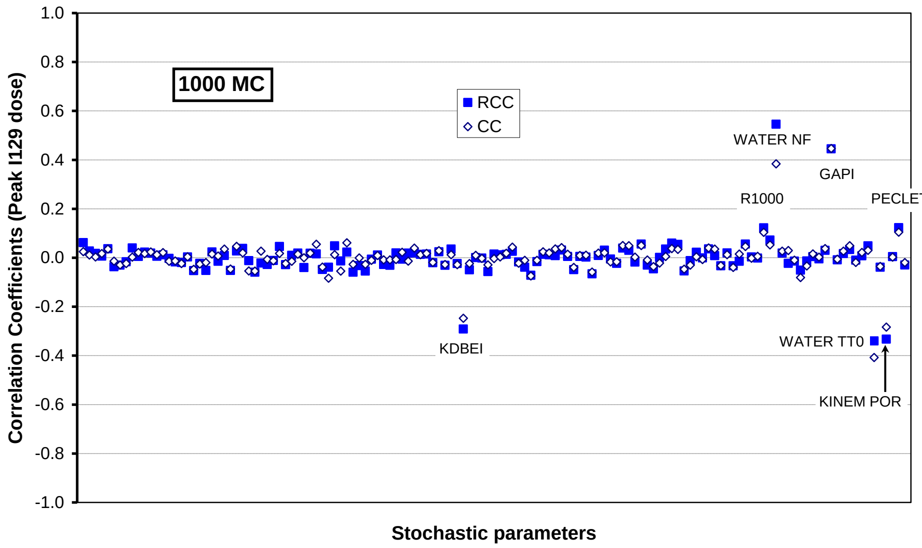


Figure 4.3 – CCs and RCCs for the peak dose due to I129 in a calculation with 1000 runs.

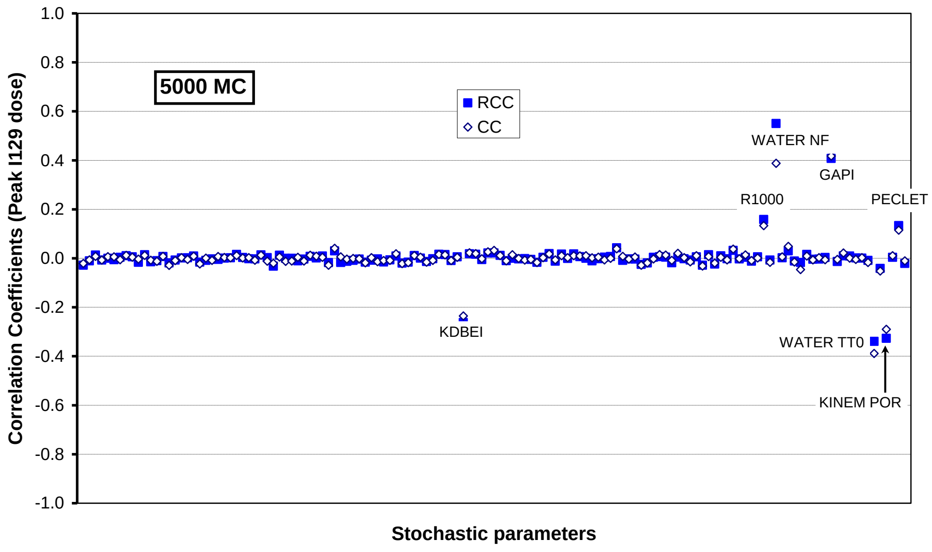


Figure 4.4 – CCs and RCCs for the peak dose due to I129 in a calculation with 5000 runs.

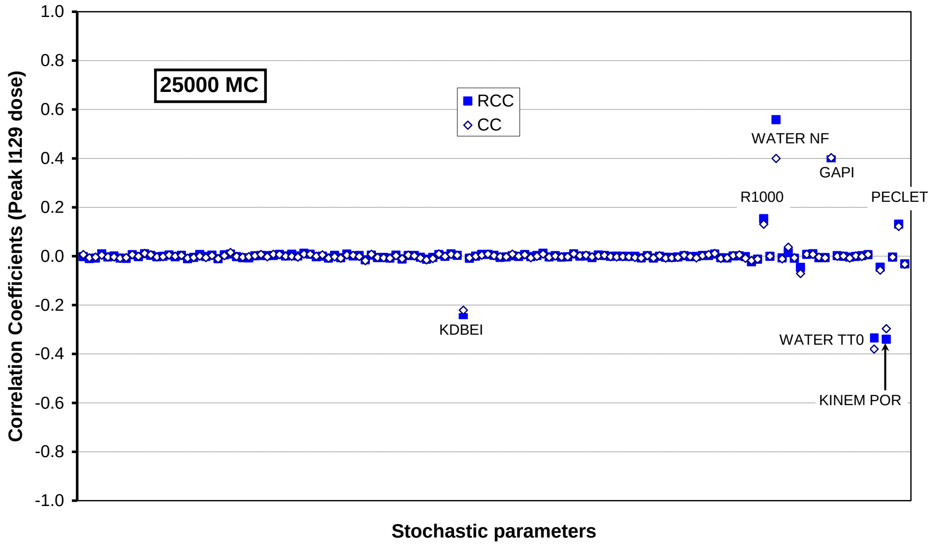


Figure 4.5 – CCs and RCCs for the peak dose due to I129 in a calculation with 25000 runs.

The increase in the number of runs allows identifying small correlations, as can be seen in Figure 4.6, where the weak correlations of MEAN and SLOPE (2 parameters that define the sequential failure of the canisters) with the peak dose due to I129 are shown. In addition, 2 parameters of the granite matrix (THICKM and FWA) are correlated with the peak dose due to I129 also, with correlation coefficients outside the band of “noise” between -0.02 and 0.02 created by the 115 parameters of the model that we know for sure that have no effect on I129 doses. The highest value (in absolute value) of a correlation coefficient in these 115 non-relevant parameters is -0.019, and corresponds to the solubility of tin (SOLSN).

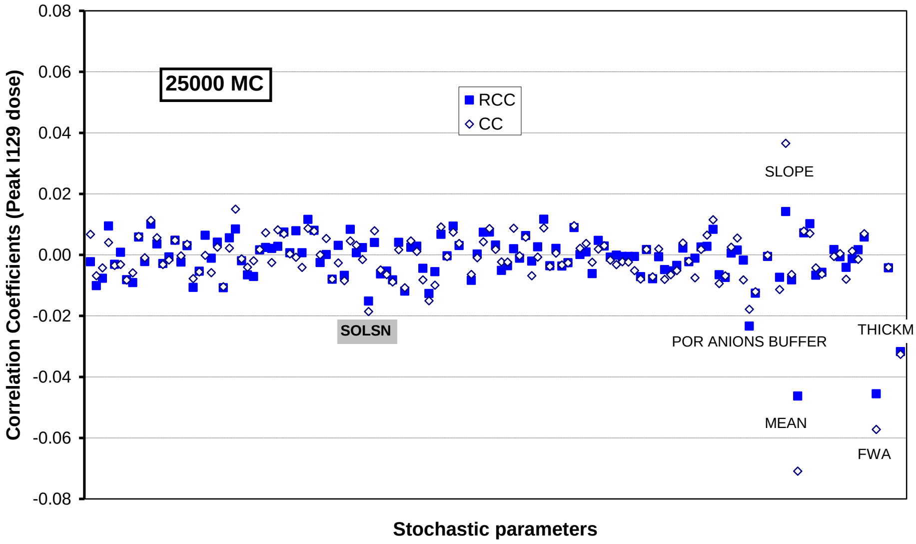


Figure 4.6 – Correlation coefficients for the peak dose due to I129 for 25000 runs (detail for correlation coefficients between -0.08 and 0.08).

4.1.1.2 Criterion for CCs and RCCs statistical significance

In Figure 4.6 it can be seen that the correlation coefficients of the 115 parameters known to have no effect on the doses due to I129 form a band around zero. The width of this band can be used to decide when a calculated CC is statistically meaningful: values outside this band are considered significant while values inside this band are not. The width of the band is defined by the "maximum value of the coefficient for non-relevant parameters", and in the case of the 25,000 runs for I129 its value is 0.019 (see Figure 4.6).

Table 4.2 shows the maximum (absolute) value of the correlation coefficients for non-relevant parameters for the peak dose due to different radionuclides in the 25,000 runs calculation. All the values are quite similar and close to 0.020, with only one exception: CCs for Cs135. For this reason, in the analyses performed in the next sections only correlation coefficients higher than 0.02 (in absolute value) are considered statistically significant and included in the tables.

Table 4.2 – Maximum absolute value of the correlation coefficients for non-relevant parameters.

	CCs	RCCs
C14	0.015	0.019
Cl36	0.020	0.020
Cs135	0.028	0.017
I129	0.019	0.015
Ni59	0.015	0.020
Pd107	0.018	0.020
Se79	0.019	0.018
Sn126	0.016	0.015
Tc99	0.016	0.019

4n	0.018	0.019
4n+1	0.017	0.016
4n+2	0.017	0.018
4n+3	0.020	0.017

The minimum value of the CCs that is statistically significant is expected to decrease when the number of runs increases. The "resolution" of the correlation analysis should improve with the number of runs, which would allow identifying weak correlations. Figure 4.7 and Figure 4.8 show the evolution with the number of runs (n) of the minimum significant value of the CCs and RCCs for the peak dose due to I129. The curves $2/\sqrt{n}$ and $4/\sqrt{n}$ are represented also, and it is observed that the empirical values are always between both curves. It can be concluded that the minimum significant value of the correlation coefficients follows the "rule of thumb":

$$\text{threshold} \;\; \alpha \;\; \frac{1}{\sqrt{n}}$$

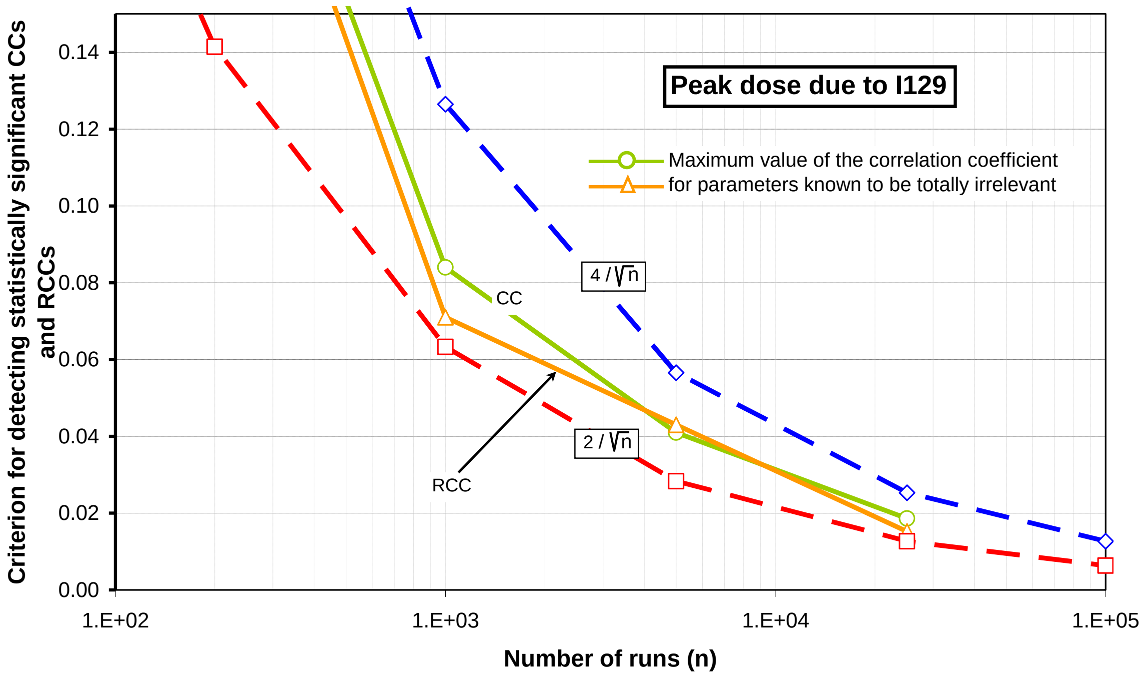


Figure 4.7 – Evolution with the number of runs of the minimum value of CCs and RCCs statistically significant for the peak dose due to I129 (output variable).

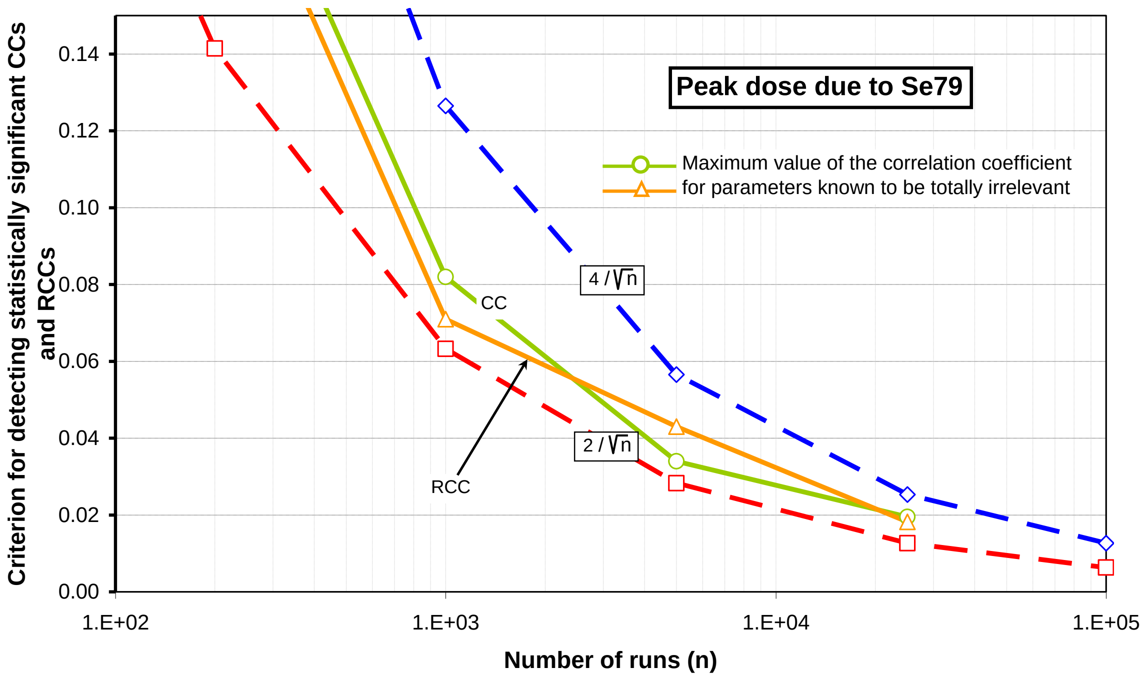
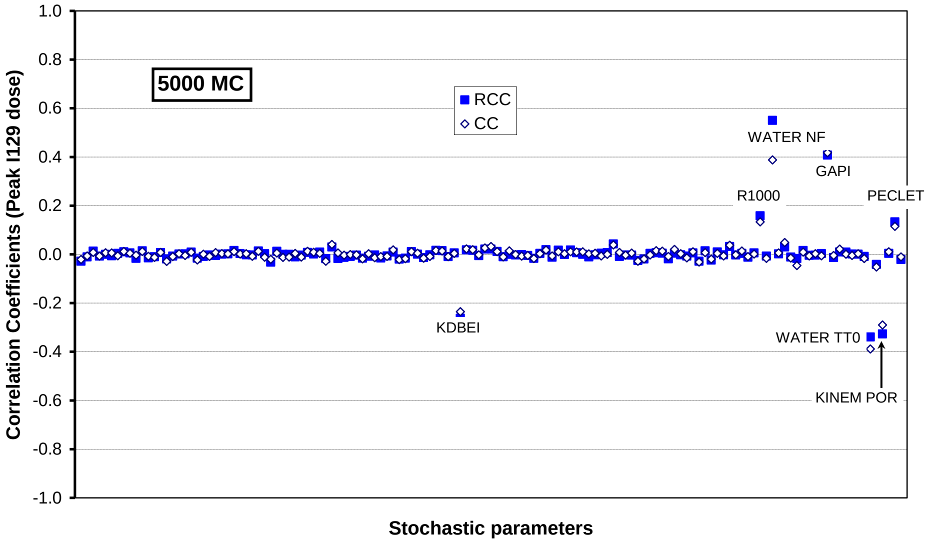


Figure 4.8 – Evolution with the number of runs of the minimum value of CCs and RCCs statistically significant for the peak dose due to Se79 (output variable).

4.1.1.3 Effect of the sampling method on the results (SRS vs. LHS.).

Stochastic calculations with 200, 1000 and 5000 runs have been done using two different sampling schemes for the random input parameters: Simple Random Sampling (SRS) and Latin Hypercube Sampling (LHS). Figure 4.9 compares the CCs and RCCs of all the random input parameters with the peak dose due to I129 in the 5000 runs calculations with SRS and LHS sampling. No significant differences between sampling methods are observed for I129 or any other radionuclide, and it can be concluded that the sampling method used (SRS or LHS) makes no difference in the correlation and regression analyses.

SRS has been adopted for the calculation with 25,000 runs used in the analyses. SRS allowed to perform several parallel stochastic calculations, each with a few thousands of runs, using different computers (or cores of a multi-core PC) and then append the runs to produce the 25,000 runs calculation. With LHS this approach is not possible and the 25,000 runs must be calculated in the same PC.



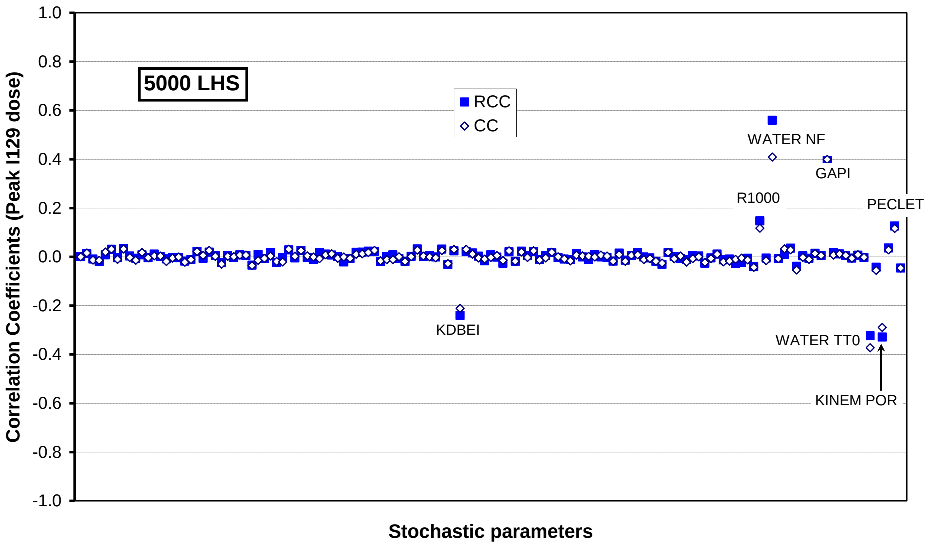


Figure 4.9 – CCs and RCCs for the dose peak I129 dose. Results for calculations with 5000 runs using SRS and LHS.

4.1.2 Analysis for random input parameters

Results of the correlation analysis for all the random input parameters and radionuclides are presented in Table 4.3 to Table 4.6. These results have been obtained in the 25,000 runs calculations, and only coefficients greater than 0.02 (in absolute value) are included, as justified in section 4.1.1.2. Element specific parameters are shown in grey cells.

Table 4.3 – Correlation analyses for total, I129, Cl36 and Cs135 peak doses.

	Total		I129		Cl36		Cs135	
	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC
Waste(3)								

R1000	0.107	0.152	0.131	0.153	0.108	0.112	0.092	0.049
METAL LIFETIME								
GAPX	0.307 (I)	0.384 (I)	0.403	0.394	0.427	0.403	0.030	
Canisters (3)								
FAIL								
MEAN	-0.056	-0.047	-0.071	-0.046	-0.108	-0.080		
SLOPE	0.026		0.037		0.056	0.036		
Near field (6)								
VC1								
POR ANIONS BUFFER		-0.022		-0.023	-0.050	-0.050		
WATER NF	0.352	0.565	0.400	0.558	0.084	0.112	0.189	0.102
SOLX	0.084(U)							
KDBEX	-0.161 (I)	-0.245 (I)	-0.221	-0.252			-0.070	-0.032
DIFFX								
Water flow (4)								
WATER TT0	-0.330	-0.340	-0.380	-0.334	-0.610	-0.590	-0.295	-0.438
KINEM POR	-0.223	-0.332	-0.297	-0.340	-0.421	-0.542		
PECLET	0.087	0.126	0.122	0.131	0.102	0.102		-0.189
PATH LENGTH								
Granite matrix (6)								
FWA	-0.066	-0.061	-0.057	-0.046	-0.075	-0.076	-0.178	-0.526
THICKM	-0.046	-0.042	-0.033	-0.032	-0.051	-0.057	-0.161	-0.365
PORMAT						-0.023		
POR ANIONS GRANITE								
MATCOR	-0.029						-0.170	-0.378
KDRX							-0.119	-0.270
R^2	0.440	0.800	0.647	0.804	0.786	0.860	0.238	0.867

Table 4.4 – Correlation analyses for C14, Ni59 and Sn126 peak doses.

	C14		Ni59		Sn126	
	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC
Waste(3)						
R1000						0.027
METAL LIFETIME		-0.026				
GAPX						
Canisters (3)						
FAIL	0.027					
MEAN		-0.020			-0.021	
SLOPE	-0.023					
Near field (6)						
VC1						
POR ANIONS BUFFER						
WATER NF	0.085	0.102	0.040	0.031	0.153	0.142
SOLX				0.031	0.024	
KDBEX	-0.086	-0.068			-0.105	-0.125
DIFFX						
Water flow (4)						

WATER TT0	-0.159	-0.566	-0.061	-0.431	-0.140	-0.407
KINEM POR	-0.056	-0.201			-0.028	-0.039
PECLET		-0.295		-0.277		-0.170
PATH LENGTH						
Granite matrix (6)						
FWA	-0.051	-0.415	-0.031	-0.473	-0.074	-0.459
THICKM	-0.051	-0.284	-0.034	-0.310	-0.059	-0.319
PORMAT						
POR ANIONS GRANITE						
MATCOR	-0.056	-0.288	-0.031	-0.334	-0.068	-0.333
KDRX	-0.039	-0.201	-0.031	-0.243	-0.067	-0.473
R^2	0.054	0.840	0.009	0.755	0.074	0.880

Table 4.5 – Correlation analyses for Tc99, Se79 and Pd107 peak doses.

	Tc99		Se79		Pd107	
	CC SRC	CC SRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC
Waste(3)						
R1000						
METAL LIFETIME						
GAPX						
Canisters (3)						
FAIL						
MEAN						
SLOPE						
Near field (6)						
VC1						
POR ANIONS BUFFER			0.028			
WATER NF	0.154	0.080	0.437	0.582	0.307	0.525
SOLX	0.090	0.044	0.492	0.627	0.538	0.574
KDBEX						
DIFFX			0.033	0.028		
Water flow (4)						
WATER TT0	-0.167	-0.428	-0.112	-0.178	-0.051	-0.165
KINEM POR			-0.041	-0.039	-0.020	
PECLET		-0.220				-0.024
PATH LENGTH						
Granite matrix (6)						
FWA	-0.088	-0.499	-0.099	-0.218	-0.068	-0.245
THICKM	-0.079	-0.349	-0.075	-0.160	-0.051	-0.164
PORMAT						
POR ANIONS GRANITE						
MATCOR	-0.071	-0.356	-0.073	-0.154	-0.051	-0.176
KDRX	-0.085	-0.376	-0.073	-0.157	-0.082	-0.257
R^2	0.086	0.879	0.475	0.887	0.403	0.817

Table 4.6 – Correlation and regression analyses for the peak doses due to the radioactive series.

	4n		4n+1		4n+2		4n+3	
	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC
Waste(3)								
R1000								
METAL LIFETIME								
GAPX								
Canisters (3)								
FAIL								
MEAN								
SLOPE								
Near field (6)								
VC1								
POR ANIONS BUFFER								
WATER NF	0.137	0.080	0.088	0.060	0.103	0.074	0.091	0.074
SOLU	0.254	0.086			0.205	0.107	0.136	0.099
SOLNP			0.125	0.075				
KDBEX								
DIFFX								
Water flow (4)								
WATER TT0	-0.124	-0.433	-0.117	-0.428	-0.097	-0.430	-0.096	-0.431
KINEM POR								
PECLET		-0.244		-0.265		-0.253		-0.253
PATH LENGTH								
Granite matrix (6)								
FWA	-0.072	-0.522	-0.061	-0.512	-0.052	-0.519	-0.050	-0.520
THICKM	-0.057	-0.350	-0.058	-0.337	-0.044	-0.348	-0.048	-0.351
PORMAT								
POR ANIONS GRANITE								
MATCOR	-0.071	-0.371	-0.056	-0.358	-0.054	-0.369	-0.053	-0.372
KDRCM								-0.030
KDRNP				-0.255				
KDRPA							-0.044	-0.046
KDRRA					-0.030			
KDRTH		-0.093	-0.046	-0.023	-0.021			
KDRU	-0.042	-0.201		-0.037		-0.236		-0.218
R^2	0.114	0.842	0.049	0.833	0.071	0.848	0.046	0.848

From the previous analyses it is possible to identify a few parameters with negligible effect on peak doses:

- VC1, PATH LENGTH and POR ANIONS GRANITE are below the threshold value of 0.02 in all the cases,
- DIFFX is slightly greater than the threshold only for Se79.

- PORMAT is slightly greater than the threshold only for Cl36.
- FAIL and METAL LIFETIME are slightly greater than the threshold only for C14.

For **solubility controlled radionuclides** (Tc99, Se79, Pd107 and the 4 radioactive series, that are controlled by long lived isotopes of Uranium and Np237):

- all the parameters related to the canister failure and waste alteration have no effect on the peak doses.
- only two parameters of the near field are important: the solubility of the radionuclide of the long lived parent of the series (U236, Np237, U238 and U235) and the water flow leaving the near field (WATER NF).
- the important parameters of the far field are those that appear in the expression of the retarded travel time in the geosphere for a sorbed species (section 4.1.4): WATER TTO, FWA, THICKM, MATCOR and KDRX.

Some of the radionuclides in the inventory have an unlimited solubility in the calculations (I129, Cl36, Cs135 and C14). Ni59 and Sn126 have a solubility limit but it is not reached in most calculations and the correlation coefficients of SOLX with the peak doses are very small or below 0.02. For these **non-solubility controlled radionuclides**:

- the parameters related to the canister failure (MEAN and SLOPE) have a small effect on the peak doses. The greatest effect is observed for Cl36.
- the inventory in the Instant Release Fraction (GAPX) is strongly correlated with the peak doses for I129 and Cl36. R1000 is an important parameter for I129, Cl36 and Cs135.
- for radionuclides with sorption on bentonite and granite (Cs135, C14, Ni59 and Sn126) the correlation coefficients of the far field parameters are much greater than for the near field parameters. Peak doses are controlled by the far field, and the role of the near field is secondary.
- for Cl36, that has no sorption on bentonite or granite, the correlation coefficients of the far field parameters are much greater than for the near field parameters also.
- only when there is sorption on bentonite but not on granite (I129) the correlation coefficients of a near field parameter (WATER NF) is greater than the correlations coefficients of the far field parameters.

Peak **total doses** are controlled by I129, and the correlation coefficients of all the parameters with the peak total dose and the peak dose due to I129 are very similar (see Table 4.3). The only exception is SOLU in the correlation analysis in values. In the 25,000 runs calculations the peak dose due to the 4n+2 series (controlled by U238) is greater than the peak dose due to I129 in 168 runs, and some of these runs produce very high doses. The maximum value of the peak dose in the 25,000 runs calculation is $5.3 \cdot 10^{-6}$ Sv/yr for I129 and $3.7 \cdot 10^{-5}$ Sv/yr for the 4n+2 series. The effect of these few runs with very high peak total dose controlled by the 4n+2 parameters is observed in the correlation analysis in values but when the transformation into ranks is done this effect disappears.

The correlation coefficients and the coefficient of determination (R^2) of the multiple regressions in ranks are much greater than in values. This is an expected result because there is a relation of monotony between the input parameters and the output variable, but this relation is not necessarily linear.

4.1.3 Transformation into logarithms of the input and output variables.

Since the random input parameters and the output variables of the model typically cover several orders of magnitude it is possible that a transformation into logarithms of both the input parameters and the result (peak dose) could be useful, in the same way that the rank transformation.

The input parameters in the model are always positive, which makes possible the transformation into logarithms. Only the peak total dose, and the peak doses due to I129, Cl36 and Se79 are non-zero in the 25,000 runs of the calculation (Table 4.1). In these 4 cases a transformation into logarithms of the peak doses and the input parameters can be done and the corresponding correlation and coefficients can be calculated.

As justified in section 4.1, since the random input parameters are independent (orthogonal) the regression coefficients are identical to the correlation coefficients. Under these conditions, the coefficients of determination (R^2) of the multiple regressions can be calculated simply as the sum of the squares of the correlation coefficients: $R^2=CC_1^2+CC_2^2+$……..

The results are presented in Table 4.7 and Table 4.8 in the grey columns “CC/SRC log” and “RCC/SRRC log”, together with the CC and RCC for comparison purposes.

Table 4.7 – Effect of the logarithmic transformation for the total peak dose and the peak dose due to I129.

	Total			I129		
	CC SRC	CC/SRC log	RCC SRRC	CC SRC	CC/SRC Log	RCC SRRC
Waste(3)						
R1000	0.107	0.165	0.152	0.131	0.168	0.153
GAPX	0.307 (I)	0.396 (I)	0.384 (I)	0.403	0.412	0.394
Canisters (3)						
MEAN	-0.056	-0.052	-0.047	-0.071	-0.053	-0.046
SLOPE	0.026			0.037		
Near field (6)						
POR ANIONS BUFFER		-0.021	-0.022		-0.023	-0.023
WATER NF	0.352	0.569	0.565	0.400	0.563	0.558
SOLX	0.084(U)					
KDBEX	-0.161 (I)	-0.232 (I)	-0.245 (I)	-0.221	-0.242	-0.252
Water flow (4)						
WATER TT0	-0.330	-0.340	-0.340	-0.380	-0.328	-0.334
KINEM POR	-0.223	-0.321	-0.332	-0.297	-0.333	-0.340

PECLET	0.087	-0.122	0.126	0.122	0.129	0.131
Granite matrix (6)						
FWA	-0.066	-0.067	-0.061	-0.057	-0.047	-0.046
THICKM	-0.046	-0.042	-0.042	-0.033	-0.029	-0.032
MATCOR	-0.029					
R^2	0.440	0.805	0.800	0.647	0.815	0.804

Table 4.8 – Effect of the logarithmic transformation for the peak dose due to Cl36 and Se79.

	Cl36			Se79		
	CC SRC	CC/SRC Log	RCC SRRC	CC SRC	CC/SRC Log	RCC/ SRRC
Waste(3)						
R1000	0.108	0.120	0.112			
GAPCL	0.427	0.397	0.403			
Canisters (3)						
MEAN	-0.108	-0.077	-0.080			
SLOPE	0.056	0.032	0.036			
Near field (6)						
POR ANIONS BUFFER	-0.050	-0.046	-0.050	0.028	0.021	
WATER NF	0.084	0.106	0.112	0.437	0.523	0.582
SOLSE				0.492	0.555	0.627
DIFFSE				0.033	0.025	0.028
Water flow (4)						
WATER TT0	-0.610	-0.580	-0.590	-0.112	-0.191	-0.178
KINEM POR	-0.421	-0.569	-0.542	-0.041	-0.035	-0.039
PECLET	0.102	0.094	0.102			
Granite matrix (6)						
FWA	-0.075	-0.075	-0.076	-0.099	-0.259	-0.218
THICKM	-0.051	-0.057	-0.057	-0.075	-0.185	-0.160
PORMAT		-0.024	-0.023			
MATCOR				-0.073	-0.176	-0.154
KDRSE				-0.073	-0.202	-0.157
R^2	0.786	0.871	0.860	0.475	0.793	0.887

The correlation coefficients calculated with the logarithmic transformation are quite similar to the coefficients calculated with the rank transformation. The values of coefficient of determination R^2 in the regression analyses increase significantly compared with the analyses in the raw values and can reach values similar to those obtained in the analysis in ranks for total, I129 and Cl36 peak doses.

Whenever possible, the transformation into logarithms of the input parameters and the output variable is recommended. This transformation increases significantly the correlation coefficients and R^2 values, becoming similar to those obtained with the transformation into ranks. The transformation into logarithms is preferred over the transformation into ranks because it provides an idea of the functional dependence of the input parameters with the output variable.

4.1.4 Analysis for derived parameters.

The analyses presented in section 4.1.1 to 4.1.3 have been done for all the random input parameters in the model, which are totally independent.

Using the experience gained in the Safety Assessment it is possible to identify some key parameters that control the doses due to the repository and are a combination of the random input parameters. In this document such parameters are called "derived parameters".

For the assessment of the Spanish repository concept in granite the following "derived parameters" have been identified:

Travel time in the geosphere (far field):

$$TT\ \ FF = \frac{WATER\ \ TT0}{1E-4}\cdot KINEM\ \ POR\ \cdot$$

$$\left[1+\left(PORMAT\cdot POR\ ANIONS\ BUFFER + (1-PORMAT)\cdot\rho\cdot MATCOR\ \ KDRX\right)\frac{THICKM\cdot FWA}{KINEM\ \ POR}\right]$$

where ρ is the density of granite solid (2630 kg/m^3), and the meaning of the other magnitudes is explained in Table 3.1. The factor POR ANIONS BUFFER only applies to C, Cl, Se, Pd, Sn and I.

For an instant injection of solute into the granite this parameter quantifies the time required by the contaminant to cross the granite formation, defined as the instant when the peak release rate is reached. This value coincides with the time period over which any instant injection to the geosphere will be spread when released to the biosphere.

Travel time in the near field:

$$TT\ NF = \frac{\pi\,(R\,ext^2 - R\,int^2)Lcan}{WATER\ \ NF}\left(\theta_B\cdot POR\ ANIONS\ BUFFER + (1-\theta_B)\cdot\rho\cdot KDBEX\right)$$

where ρ is the density of bentonite solid (2700 kg/m^3), θ_B the total porosity of the bentonite (0.407), Rint the radius of the canister, Rext the radius of the disposal gallery and Lcan the length of gallery corresponding to one canister and the meaning of the other magnitudes is explained in Table 3.1. The factor POR ANIONS BUFFER only applies to C, Cl, Se and I.

For an instant release from the waste of a highly soluble element this parameter quantifies the time period over which the releases from the bentonite will be spread. This value is much higher than the time required by the contaminant to cross the bentonite (reach the peak release rate).

Decay during the transport:

$$DecayT = \exp\left(-\frac{0.693 * TT\ FF}{T_{1/2}}\right)$$

where TT FF is the travel time in the far field defined in this section and $T_{1/2}$ the radionuclide half-life.

This parameter quantifies the radioactive decay during transport. To take into account the transport in the near field a value in the order of 0.01*TT_NF could be added to TT FF but the sum would be nearly identical to TT FF.

Mass flow from near field:

$$MF\ NF = SOLX \cdot WATER\ NF$$

where SOLX is the solubility limit of element X and WATER NF is the groundwater flow in the granite of the near field.

For solubility controlled elements this parameter quantifies the long term flux of solute leaving the near field, neglecting radioactive decay.

4.1.4.1 Correlation analyses.

Correlation analyses for all the “derived parameters” and relevant radionuclides have been done, and the results are presented in Table 4.9 to Table 4.12. The grey cells of the tables contain the correlation coefficients for the derived parameters that have been identified as potentially useful. The correlation coefficients of the random input parameters are presented also (white cells) to facilitate the comparison of the results.

Table 4.9 – Correlation and regression analyses for total, I129, Cl36 and Cs135 peak doses.

	Total		I129		Cl36		Cs135	
	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC
Waste(3)								
R1000	0.107	0.152	0.131	0.153	0.108	0.112	0.092	0.049
GAPX	0.307 (I)	0.384 (I)	0.403	0.394	0.427	0.403	0.030	
Canisters (3)								
MEAN	-0.056	-0.047	-0.071	-0.046	-0.108	-0.080		
SLOPE	0.026		0.037		0.056	0.036		
Near field (6)								
POR ANIONS BUFFER		-0.022		-0.023	-0.050	-0.050		
WATER NF	0.352	0.565	0.400	0.558	0.084	0.112	0.189	0.102
SOLX	0.084(U)							
KDBEX	-0.161	-0.245 (I)	-0.221	-0.252			-0.070	-0.032
TT NF	-0.399	-0.618 (I)	-0.495	-0.615	-0.166	-0.119	-0.151	-0.107
1/TT NF	0.214	0.618 (I)	0.257	0.615	0.073	0.119	0.185	0.107
Water flow (4)								
WATER TT0	-0.330	-0.340	-0.380	-0.334	-0.610	-0.590	-0.295	-0.438
KINEM POR	-0.223	-0.332	-0.297	-0.340	-0.421	-0.542		
PECLET	0.087	0.126	0.122	0.131	0.102	0.102		-0.189
Granite matrix (6)								
FWA	-0.066	-0.061	-0.057	-0.046	-0.075	-0.076	-0.178	-0.526

THICKM	-0.046	-0.042	-0.033	-0.032	-0.051	-0.057	-0.161	-0.365
PORMAT						-0.023		
MATCOR	-0.029						-0.170	-0.378
KDRX							-0.119	-0.270
TT FF	-0.379	-0.504 (I)	-0.465	-0.502	-0.687	-0.862	-0.125	-0.958
1/TT FF	0.300	0.504 (I)	0.333	0.502	0.557	0.862	0.523	0.958
DecayT	0.378	0.504 (I)	0.466	0.502	0.753	0.862	0.685	0.958
TT NF+TT FF	-0.549	-0.863 (I)	-0.679	-0.862	-0.699	-0.878	-0.132	-0.955
1/(TT NF+TT FF)	0.579	0.863 (I)	0.667	0.862	0.623	0.878	0.928	0.955
GAP/(TT NF+TT FF)	0.662	0.948 (I)	0.780	0.952	0.720	0.964	0.921	0.919
DecayT/(TT NF+TT FF)	0.579	0.862 (I)	0.667	0.862	0.613	0.877	0.906	0.965

Table 4.10 – Correlation and regression analyses for C14, Ni59 and Sn126 peak doses.

	C14		Ni59		Sn126	
	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC
Waste(3)						
R1000						0.027
METAL LIFETIME		-0.026				
Canisters (3)						
FAIL	0.027					
MEAN		-0.020			-0.021	
SLOPE	-0.023					
Near field (6)						
WATER NF	0.085	0.102	0.040	0.031	0.153	0.142
SOLX				0.031	0.024	
KDBEX	-0.086	-0.068			-0.105	-0.125
TT NF	-0.062	-0.122	-0.021	-0.034	-0.070	-0.190
1/TT NF	0.116	0.122	0.026	0.034	0.290	0.190
MF NF			0.029	0.040	0.114	0.102
Water flow (4)						
WATER TT0	-0.159	-0.566	-0.061	-0.431	-0.140	-0.407
KINEM POR	-0.056	-0.201			-0.028	-0.039
PECLET		-0.295		-0.277		-0.170
Granite matrix (6)						
FWA	-0.051	-0.415	-0.031	-0.473	-0.074	-0.459
THICKM	-0.051	-0.284	-0.034	-0.310	-0.059	-0.319
MATCOR	-0.056	-0.288	-0.031	-0.334	-0.068	-0.333
KDRX	-0.039	-0.201	-0.031	-0.243	-0.067	-0.473
TT FF	-0.054	-0.932	-0.017	-0.880	-0.043	-0.952
1/TT FF	0.467	0.932	0.485	0.880	0.296	0.952
DecayT	0.486	0.932	0.587	0.894	0.337	0.952
TT NF+TT FF	-0.077	-0.721	-0.017	-0.872	-0.079	-0.463
1/(TT NF+TT FF)	0.681	0.721	0.428	0.872	0.689	0.463
DecayT/(TT NF+TT FF)	0.871	0.936	0.855	0.894	0.890	0.972
MF NF*DecayT			0.800	0.894	0.419	0.961

Table 4.11 – Correlation and regression analyses for Tc99, Se79 and Pd107 peak doses.

	Tc99		Se79		Pd107	
			CC SRC	RCC SRRC	CC SRC	RCC SRRC
Near field (6)						
POR ANIONS BUFFER			0.028			
WATER NF	0.154	0.080	0.437	0.582	0.307	0.525
SOLX	0.090	0.044	0.492	0.627	0.538	0.574
DIFFX			0.033	0.028		
TT NF	-0.076	-0.080	-0.223	-0.511	-0.126	-0.410
1/TT NF	0.131	0.080	0.178	0.511	0.144	0.410
MF NF	0.211	0.091	0.837	0.863	0.847	0.787
Water flow (4)						
WATER TT0	-0.167	-0.428	-0.112	-0.178	-0.051	-0.165
KINEM POR			-0.041	-0.039	-0.020	
PECLET		-0.220				-0.024
Granite matrix (6)						
FWA	-0.088	-0.499	-0.099	-0.218	-0.068	-0.245
THICKM	-0.079	-0.349	-0.075	-0.160	-0.051	-0.164
MATCOR	-0.071	-0.356	-0.073	-0.154	-0.051	-0.176
KDRX	-0.085	-0.376	-0.073	-0.157	-0.082	-0.257
TT FF	-0.052	-0.959	-0.120	-0.395	-0.081	-0.456
1/TT FF	0.487	0.959	0.110	0.395	0.057	0.456
DecayT	0.526	0.959	0.220	0.395	0.156	0.456
TT NF+TT FF	-0.066	-0.797	-0.137	-0.540	-0.122	-0.653
1/(TT NF+TT FF)	0.584	0.797	0.277	0.540	0.211	0.653
DecayT/(TT NF+TT FF)	0.768	0.962	0.267	0.511	0.205	0.641
MF NF*DecayT	0.983	0.963	0.974	0.995	0.988	0.980

Table 4.12 – Correlation and regression analyses for the peak doses due to the radioactive series.

	4n		4n+1		4n+2		4n+3	
	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC
Near field (6)								
WATER NF	0.137	0.080	0.088	0.060	0.103	0.074	0.091	0.074
SOLU	0.254	0.086			0.205	0.107	0.136	0.099
SOLNP			0.125	0.075				
TT NF	-0.072	-0.080	-0.051	-0.061	-0.053	-0.074	-0.047	-0.075
1/TT NF	0.112	0.080	0.077	0.061	0.083	0.074	0.075	0.075
MF NF	0.396	0.118	0.165	0.095	0.324	0.130	0.232	0.124
Water flow (4)								
WATER TT0	-0.124	-0.433	-0.117	-0.428	-0.097	-0.430	-0.096	-0.431
KINEM POR								
PECLET		-0.244		-0.265		-0.253		-0.253
Granite matrix (6)								

FWA	-0.072	-0.522	-0.061	-0.512	-0.052	-0.519	-0.050	-0.520
THICKM	-0.057	-0.350	-0.058	-0.337	-0.044	-0.348	-0.048	-0.351
MATCOR	-0.071	-0.371	-0.056	-0.358	-0.054	-0.369	-0.053	-0.372
KDRCM								-0.030
KDRNP				-0.255				
KDRPA							-0.044	-0.046
KDRRA					-0.030			
KDRTH		-0.093	-0.046	-0.023	-0.021			
KDRU	-0.042	-0.201		-0.037		-0.236		-0.218
TT FF	-0.055	-0.923	-0.036	-0.927	-0.040	-0.931	-0.032	-0.928
1/TT FF	0.167	0.923	0.221	0.927	0.097	0.931	0.137	0.928
DecayT			0.263	0.927				
TT NF+TT FF	-0.066	-0.808	-0.050	-0.688	-0.048	-0.815	-0.044	-0.812
1/(TT NF+TT FF)	0.388	0.808	0.296	-0.688	0.265	0.815	0.259	0.812
DecayT/(TT NF+TT FF)			0.332	0.929				
MF NF*DecayT			0.722	0.933				

Results in Table 4.9 to Table 4.12 show that “derived parameters” are much more correlated with the peak dose than the original random input parameters. The sensitivity analysis confirms that the “derived parameters” are the parameters that really control the behaviour of the radionuclide in the repository and consequently, their doses.

“Derived parameters” should be identified at the beginning of the sensitivity analysis and treated in the same way that the original random input parameters through the analysis, keeping in mind that these “derived parameters” are correlated with some input parameters. This is the approach followed in the sections 4.2 to 4.7, where the analyses are performed for all the random input parameters and derived parameters.

4.1.4.2 Regression analysis

The correlation coefficients obtained in section 4.1.4.1 can be used to make regression analyses for different sets of parameters. The parameters considered in a given analysis are independent (not correlated), which allows to calculate the coefficient of determination R^2 as the sum of the squares of the correlation coefficients: $R^2 = CC_1^2 + CC_2^2 + \ldots\ldots$

The results of the regression analyses are shown in Table 4.13. For the peak total dose, the peak dose due to I129, the peak dose due to Se79 and the peak dose due to 4n+2 series the results of 4 different “regression analyses” are presented. Each “regression analysis” is presented in three columns: a column of wedges that identify the parameters included in the analysis, the CCs (equal to the SRCs) and the RCCs (equal to the SRRCs).

Table 4.13 – Regression analysis for random input parameters and derived parameters identified as relevant in the correlation analysis

TOTAL		CC	RCC		CC	RCC		CC	RCC		CC	RCC
GAPI	▶	0.307	0.384	▶	0.307	0.384						
R1000	▶	0.107	0.152	▶	0.107	0.152	▶	0.109	0.148			
PECLET	▶	-0.087	0.126	▶	-0.087	0.126	▶	0.088	0.124			

TT NF	►	-0.399	-0.618									
TT FF	►	-0.379	-0.504									
1/(TT NF+TT FF)				►	0.579	0.863						
GAP/(TT NF+TT FF							►	0.662	0.948	►	0.662	0.948
R^2		0.416	0.822		0.449	0.931		0.458	0.936		0.438	0.899
R^2 with all the random input parameters											0.440	0.800

I129		CC	RCC		CC	RCC		CC	RCC		CC	RCC
GAPI	▶	0.403	0.394	▶	0.403	0.394						
R1000	▶	0.131	0.153	▶	0.131	0.153	▶	0.131	0.153			
PECLET	▶	0.122	0.131	▶	0.122	0.131	▶	0.122	0.131			
TT NF	▶	-0.495	-0.615									
TT FF	▶	-0.465	-0.502									
1/(TT NF+TT FF)				▶	0.667	0.862						
GAP/(TT NF+TT FF)							▶	0.780	0.952	▶	0.780	0.952
R^2		0.656	0.826		0.639	0.939		0.640	0.947		0.608	0.906
R^2 with all the random input parameters											0.647	0.804

Se79		CC	RCC		CC	RCC		CC	RCC		CC	RCC
SOLSE	▶	0.492	0.627	▶	0.492	0.627						
WATER NF	▶	0.437	0.582	▶	0.437	0.582						
TT FF	▶	-0.120	-0.395									
DecayT				▶	0.220	0.395	▶	0.220	0.395			
SOLSE*WATER NF							▶	0.837	0.863			
Global										▶	0.974	0.995
R^2		0.447	0.888		0.481	0.898		0.749	0.901		0.949	0.990
R^2 with all the random input parameters											0.475	0.887

4n+2 series		CC	RCC		CC	RCC		CC	RCC		CC	RCC
SOLU	▶	0.205	0.107									
WATER NF	▶	0.103	0.074									
PECLET	▶	--	-0.253	▶	--	-0.253	▶	--	-0.253			
TT FF	▶	-0.040	-0.931	▶	-0.040	-0.931	▶	-0.040	-0.931	▶	-0.040	-0.931
SOLU*WATER NF				▶	0.324	0.130						
R^2		0.054	0.948		0.107	0.953		0.002	0.931		0.002	0.867
R^2 with all the random input parameters											0.071	0.848

Table 4.13 shows that the variance of the peak doses (total, I129, Se79 and 4n+2 series) can be explained by the variance of a few (or even just one) “derived parameters”. This is specially valid when the transformation into ranks is done.

It is noteworthy that the variance of the peak dose due to Se79 is totally controlled by a single “derived parameter”:

$$\text{Global}(\text{Se}79) = \text{SOLSE} * \text{WATER NF} * \exp\left(- \frac{0.693 * \text{TT FF}(\text{Se})}{T_{1/2}(\text{Se}79)} \right)$$

that is a function of the Solubility of the Selenium, the water flow leaving the near field (WATER NF), the travel time of Selenium in the far field (TT FF(Se)) and the half-life of Se79.

4.2 Time evolution of the correlation and regression coefficients

4.2.1 Total dose

Figure 4.10 and Figure 4.11 show the results of the correlation and regression analyses for the time dependent total dose in the stochastic calculation with 25,000 runs. All the random input parameters whose coefficient is greater than 0.1 (in absolute value) at least in one instant are included in the figures.

Since the random input parameters are independent (orthogonal), at all times the CCs and SRCs are identical to the RCCs and SRRCs. The relevant parameters are the same in the analyses in values and ranks with 2 exceptions: SOLU appears in the analyses in values but not in ranks while PECLET appears in the analyses in ranks but not in values.

In Figure 4.11 around 1,000 and 10,000 years the curves have strange jumps. The reason is that at those instants the dose becomes non-zero in a great number of runs. This result is related to the code used in the calculations and will be further investigated. In any case doses are very small up to 10,000 years.

In the calculations the minimum duration of the canisters is 20,000 years. Only a few canisters (FAIL) are assumed to present a defect that causes their failure in about a hundred years. As a consequence, FAIL is relevant at the beginning of the calculations but becomes irrelevant when many more canisters have failed.

Doses are controlled initially by the I129 in the gap pellet-cladding (GAPI), that is released immediately when the canister fails. After 200,000 years total doses are controlled by the I129 released from the waste at a slow rate (R1000) with a significant contribution from Se79 (SOLSE). In the very long term doses due to 4n+2 series (U238) are quite high is some runs and SOLU is relevant in the analysis in raw values but not in the analysis in ranks.

These figures show the parameters that control the peak dose at each time, but must be used with care. The UO_2 matrix alteration rate (R1000) seems to be far more important than the Instant Release Fraction of iodine (GAPI), but I129 peak dose is controlled by GAPI. When the I129 inventory in the gap has been released from the formation, ulterior I129 doses are roughly proportional to R1000.

Figure 4.10 and especially Figure 4.11 present a significant decrease in the value of the coefficient of determination (R^2) between 100,000 and 400,000 thousand years. The reasons for such drop are analysed in the next section and it was concluded that it is a consequence of the change of sign of the regression coefficients for several parameters vs. the dose due to the I129 released from the gap (see section 4.2.2).

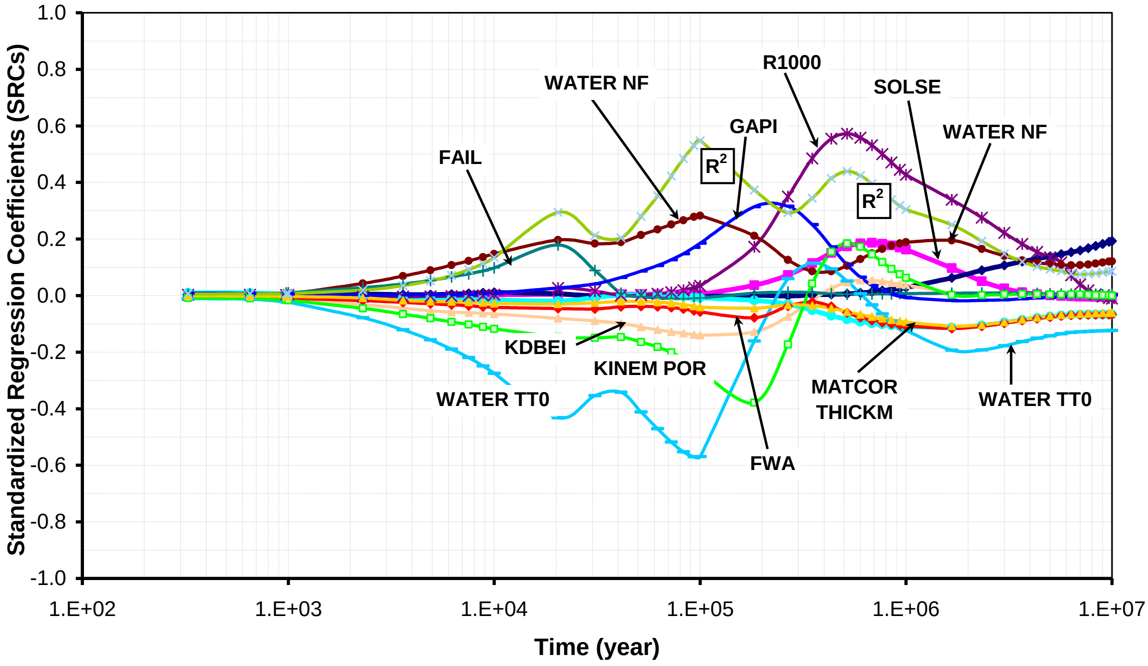


Figure 4.10 – Time evolution of the Standardized Regression Coefficients (SRCs) and the coefficient of determination (R^2) for the total dose.

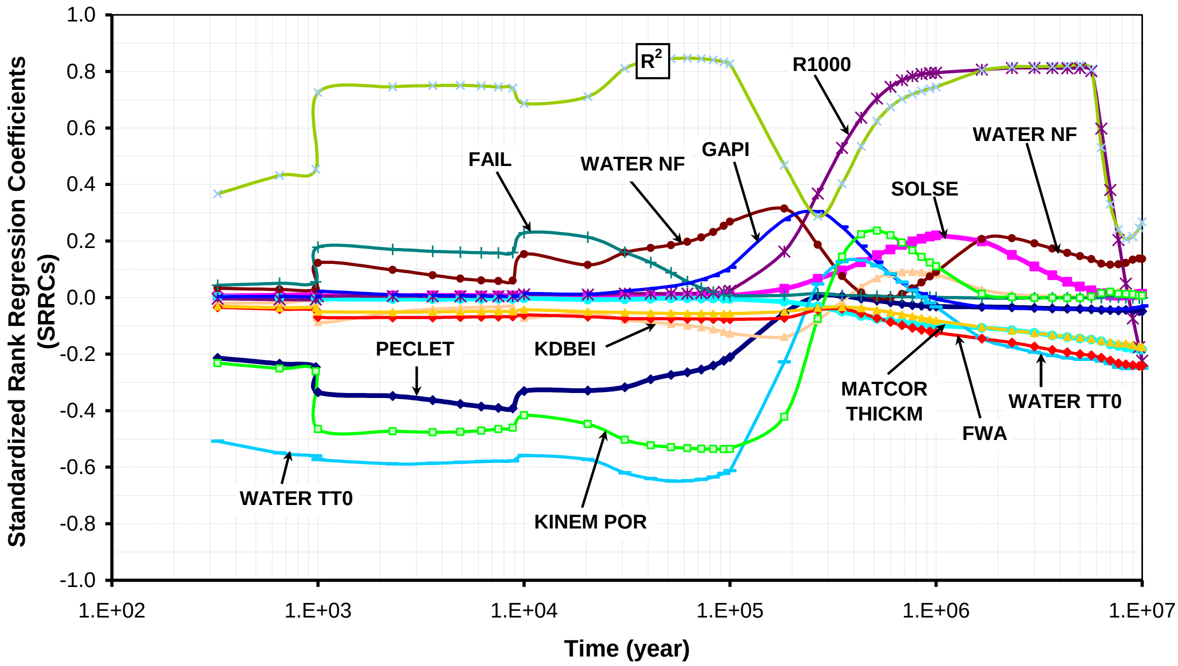


Figure 4.11 – Time evolution of the Standardized Rank Regression Coefficients (SRRCs) and the coefficient of determination (R^2) for the total dose.

4.2.2 Dose due to I129

Figure 4.12 shows the time evolution of the SRCs and R^2 for the 20 independent random input parameters relevant for iodine vs. I129 dose due to the total inventory (gap plus UO_2 matrix). The results have been obtained for a 1000 runs calculation without early failures of canisters, and as a consequence, releases and doses start after 20,000 years.

Figure 4.12 presents a significant decrease in the value of the coefficient of determination (R^2) between 100,000 and 400,000 thousand years, similar to the decrease observed for the total dose (Figure 4.11). Since doses are controlled by I129 during that time period (see Figure 4.1), this result was expected.

In order to identify the origin of this drop in R^2, independent regression analyses were done for the doses due to the I129 in the gap (Figure 4.13) and the I129 in the UO_2 matrix (Figure 4.14). For the gap inventory R^2 presents a drop between 100,000 and 400,000 thousand years, similar to the decrease observed for the total dose. By the contrary, for the inventory in the UO_2 the coefficient of determination (R^2) is nearly flat during that time period.

On the basis of the previous results it can be concluded that the significant decrease in the value of the coefficient of determination (R^2) between 100,000 and 400,000 thousand years observed for the total dose is due to a similar drop in R^2 for the dose due to the I129 in the gap.

As shown in Figure 4.13, between 100,000 and 400,000 years the SRRCs of several parameters that affect the dose due to the I129 in the gap change their sign and get close to cero. Since $R^2=SRRC_1^2+ SRRC_2^2+\ldots$ for independent random parameters, the resulting value of the coefficient of determination will drop in the transient period.

It seems surprising that the SRRC of variable passes from a high positive value to a very negative value when the time passes (or vice versa), but there is an explanation:

- For the I129 in the gap, release and mobilisation are immediate after canister failure (due to its high solubility).
- Initially, high doses are correlated with a rapid transport through the barriers, that is favoured by high values of WATER NF and low values of WATER TT0.
- But if the contaminant moves very quickly, all the inventory will be released to the biosphere in a short time period (in the order of hundreds of thousands years) and doses will be negligible thereafter.
- Long term doses are controlled by the runs with slower transport through the barriers, that correspond to low values of WATER NF and high values of WATER TT0".

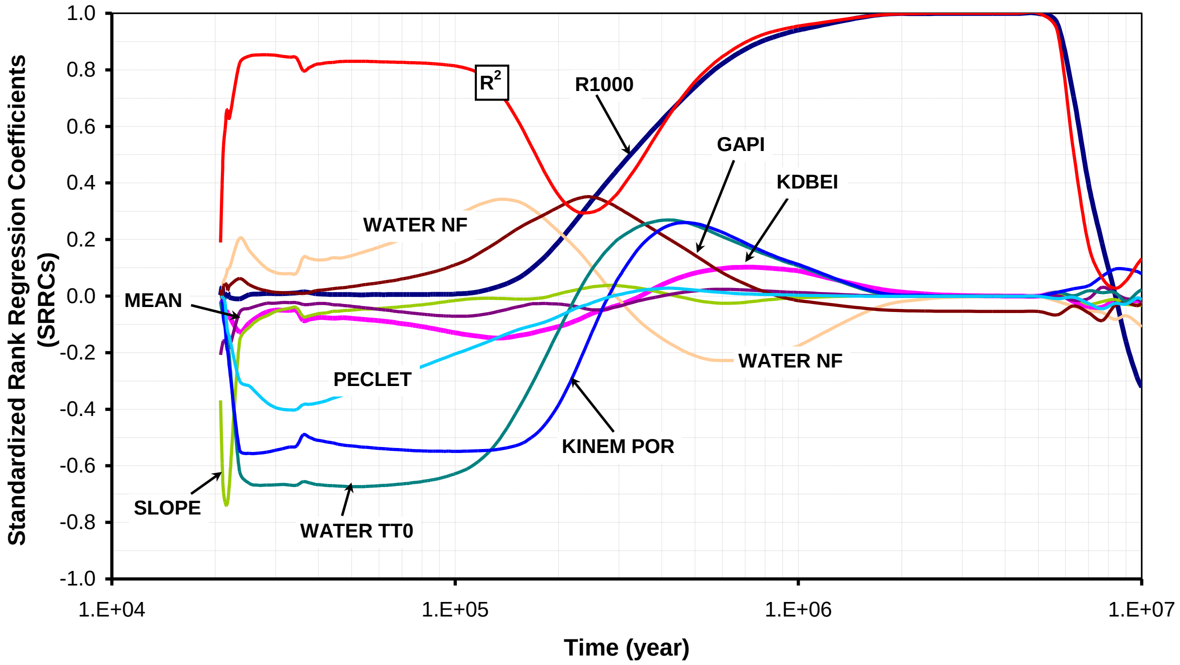


Figure 4.12 – Time evolution of the Standardized Regression Coefficients (SRCs) and the coefficient of determination (R^2) for the I129 dose due to the total inventory of I129.

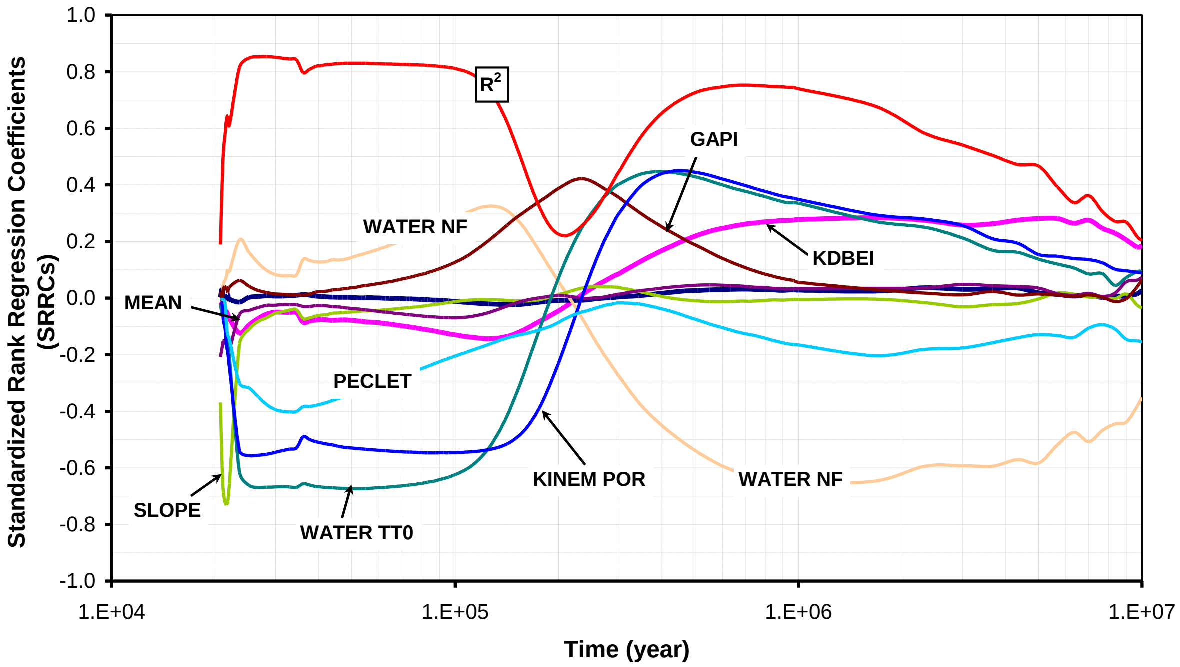


Figure 4.13 – Time evolution of the Standardized Regression Coefficients (SRCs) and the coefficient of determination (R^2) for the I129 dose due to the inventory in the gap.

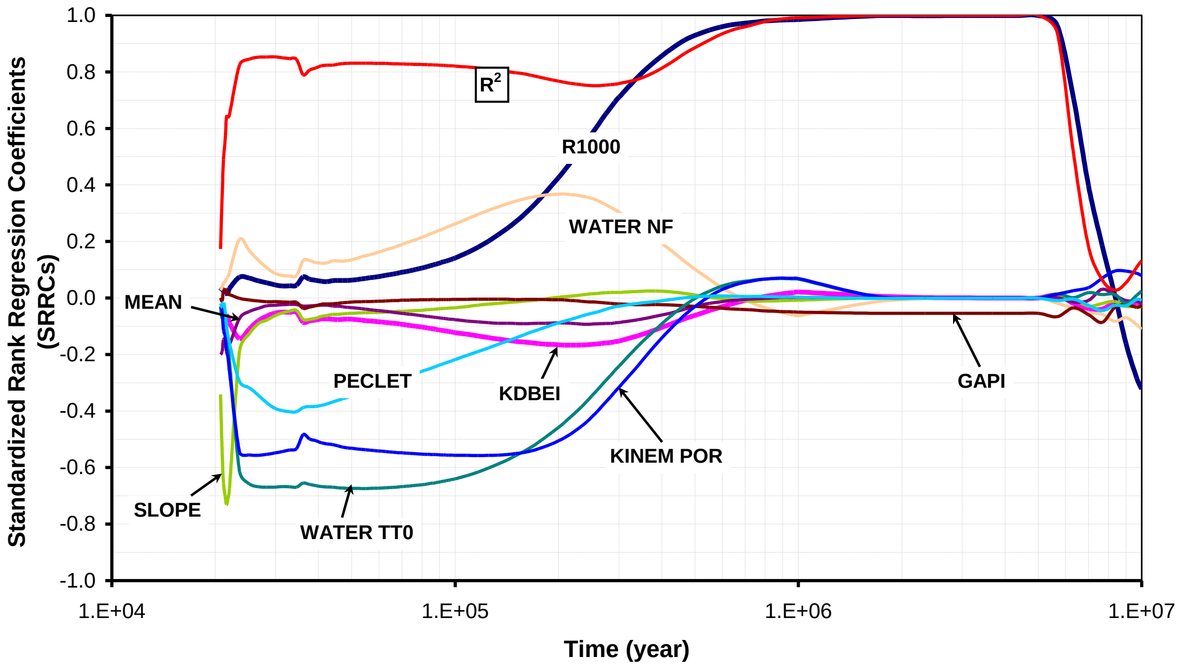


Figure 4.14 – Time evolution of the Standardized Regression Coefficients (SRCs) and the coefficient of determination (R^2) for the I129 dose due to the inventory in the UO_2.

4.2.3 Derived parameters

The derived parameters identified in section 4.1.4 were found useful in the correlation and regression analyses for the peak doses. In this section these derived parameters are used in the time dependent regression analyses.

The tests have been done for the dose due to I129 in the 1000 runs calculation without early failures of canisters used in previous section 4.2.2. A set of 5 independent input parameters was used: 3 random (PECLET, R1000 and GAPI) and 2 derived (travel times in the near field and the far field: TT NF and TT FF).

Figure 4.15 presents the time evolution of the SRRCs and R^2 for the set of 5 parameters. In addition, R^2 for the original set of 20 random input parameters is presented. It can be observed that, at any instant, the set of 5 parameters explains a similar fraction of the variance of the I129 compared with the initial set of 20 parameters, and between 25,000 and 150,000 years even a higher fraction.

Figure 4.16 presents the time evolution of the SRCs and R^2 of the set of 5 input parameters, as well as R^2 for the original set of 20 random input parameters for regression analyses in values. In this case both sets of parameters explain a similar fraction of the variance of I129 in the long term, but at earlier times there are large differences.

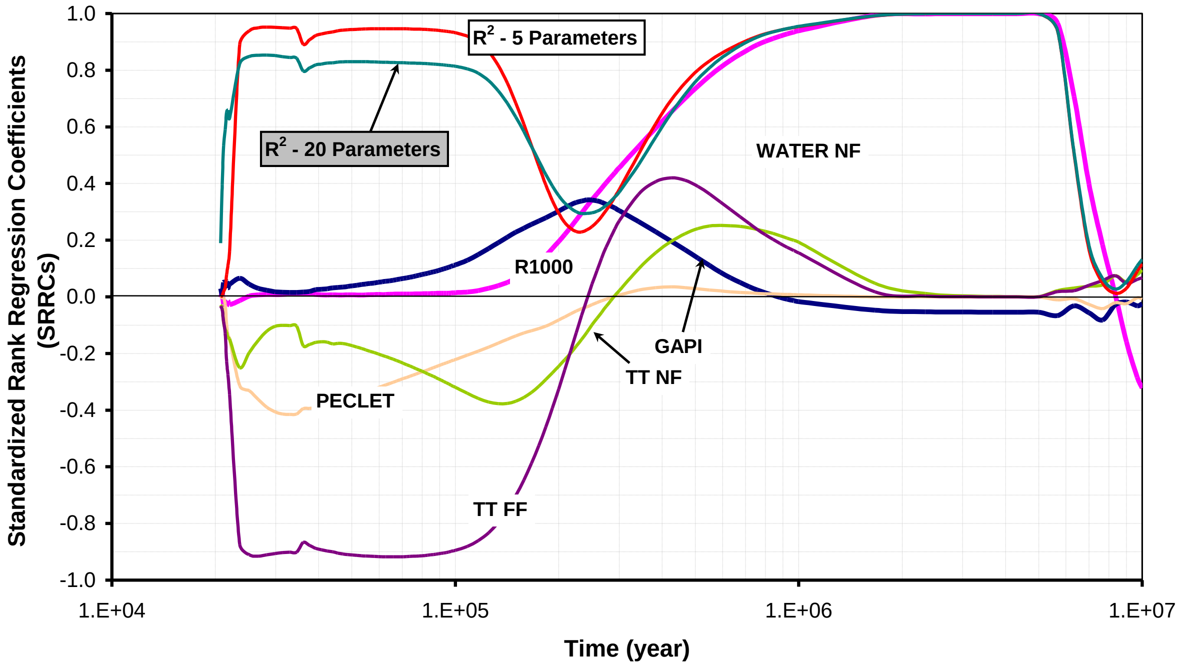


Figure 4.15 – Time evolution of the Standardized Rank Regression Coefficients (SRRCs) and the coefficient of determination (R^2) for a set of 5 parameters vs. I129 doses.

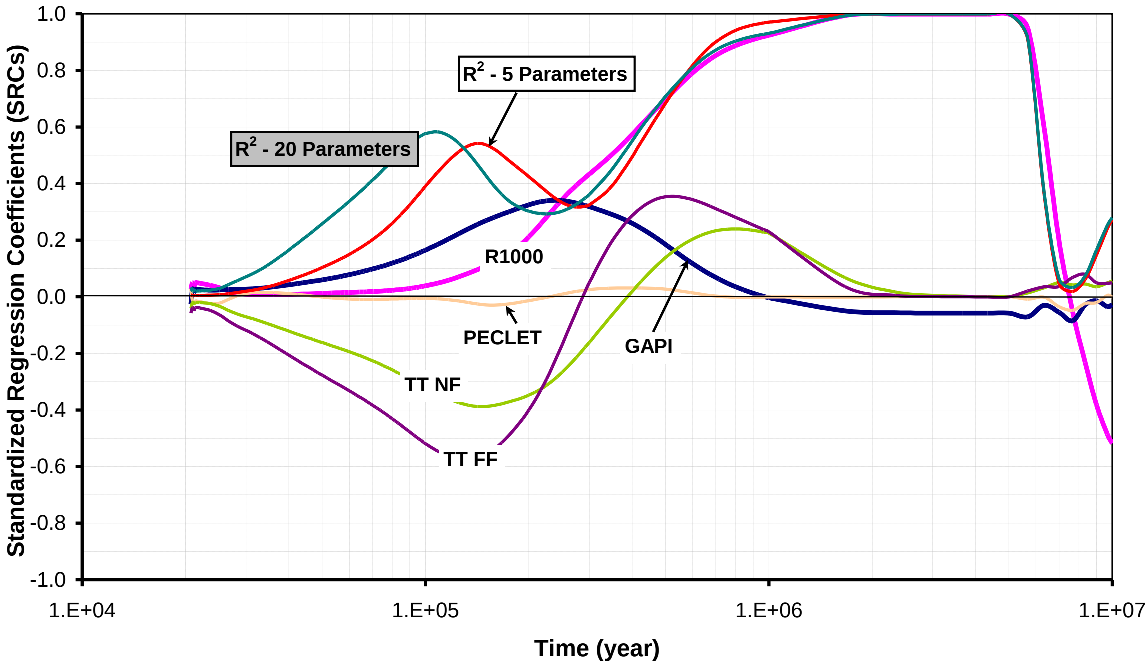


Figure 4.16 – Time evolution of the Standardized Regression Coefficients (SRCs) and the coefficient of determination (R^2) for a set of 5 parameters vs. I129 doses.

The results obtained in this section show that the derived parameters identified in section 4.1.4 are useful in time dependent regression analyses also.

4.3 Monte Carlo filtering (Smirnov, Mann-Whitney and t tests)

Monte Carlo filtering is based on dividing the output sample in two or more subsets according to some criterion (achievement of a given condition, exceeding a threshold, etc.) and testing if the inputs associated to those subsets are different or not. This can be done graphically, as in Figure 4.17 to Figure 4.21, but it is more practical to follow a numerical approach using some statistics and their associated tests, such as the two-sample Smirnov, Mann-Whitney and t tests.

The two-sample Smirnov test is based on comparing empirical cumulative distribution functions, the Mann-Whitney is based on ranks and the two-sample t-test is based on the sampling distribution of the mean of normal variables.

Monte Carlo filtering has been tested with the peak doses of different radionuclides. The output variable (peak dose) sample is divided in two sub-samples and the statistics test whether the inputs associated to those subsets are different or not.

Different criteria are used to generate the 2 subsets:

- divide the runs in the 10% producing highest doses and the remainder 90% (10%-90%),
- divide the runs in the 10% producing lowest doses and the remainder 90% (90%-10%)
- create a subset with the runs that cause a peak dose equal to zero and another with the runs that produce a non-zero peak dose (zero vs. non zero).

4.3.1 Graphic results

Figure 4.17 to Figure 4.21 present the cumulative distribution functions for several input parameters and different subsets of the 25,000 runs calculation defined on the basis of the peak dose due to C14. If a parameter is non important, the different curves in a given figure should be nearly overlapped, as can be seen in Figure 4.21 for DIFFCS, a parameter that we know for sure that can have no effect on the doses due to C14. The greater the effect of a parameter on the peak dose due to C14, more differences will be observed between the cumulative distribution functions for the different subsets represented, as can be observed comparing Figure 4.17 to Figure 4.21.

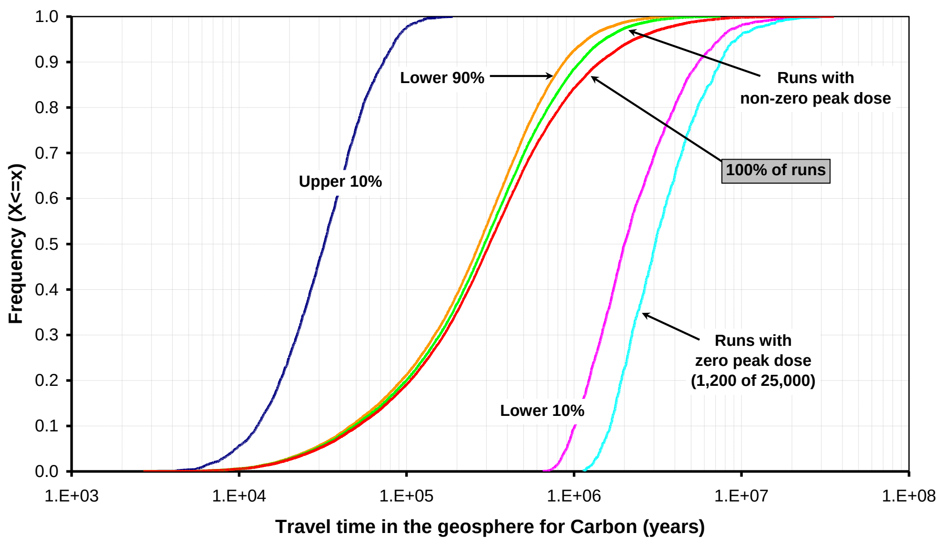


Figure 4.17 – Cumulative distribution functions of the travel time in the geosphere (TT FF) for Carbon for different subsets of the 25,000 run calculation defined on the basis of C14 peak dose.

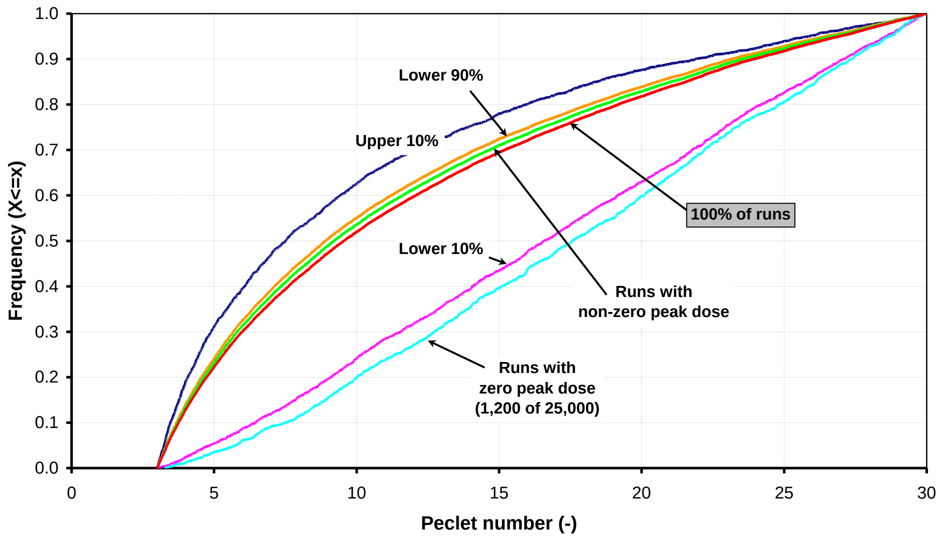


Figure 4.18 – Cumulative distribution functions of the Peclet number (PECLET) for different subsets of the 25,000 run calculation defined on the basis of peak C14 dose.

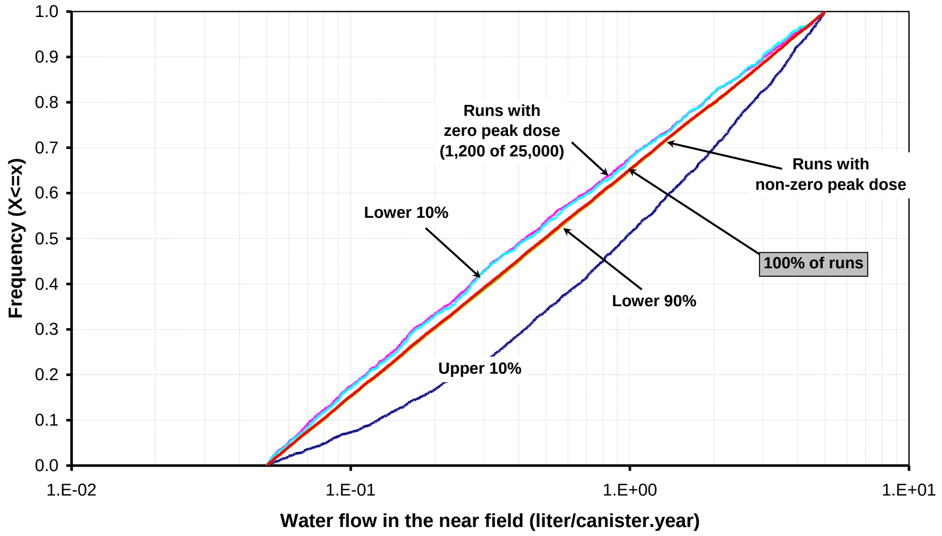


Figure 4.19 – Cumulative distribution functions of the water flow in the near field (WATER NF) for different subsets of the 25,000 run calculation defined on the basis of peak C14 dose.

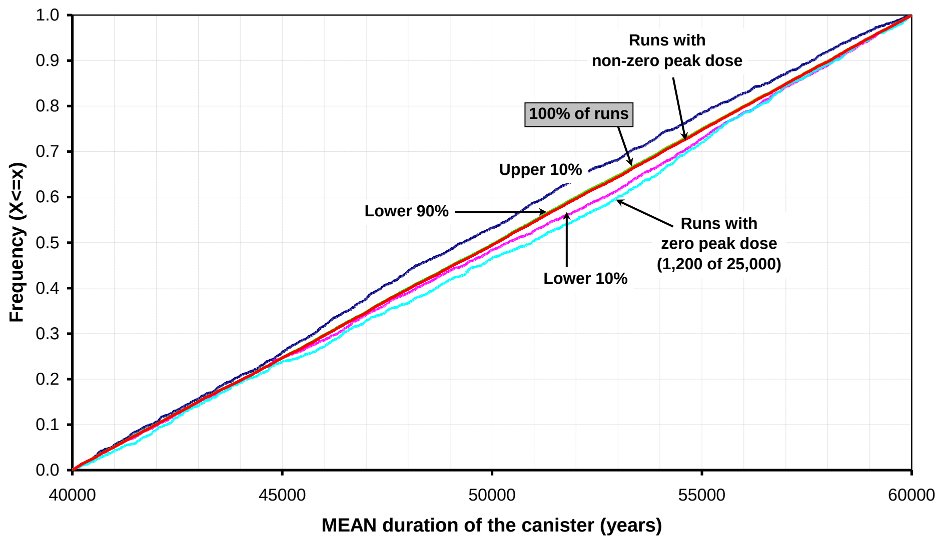


Figure 4.20 – Cumulative distribution functions of the mean duration of the canisters (MEAN) for different subsets of the 25,000 run calculation defined on the basis of peak C14 dose.

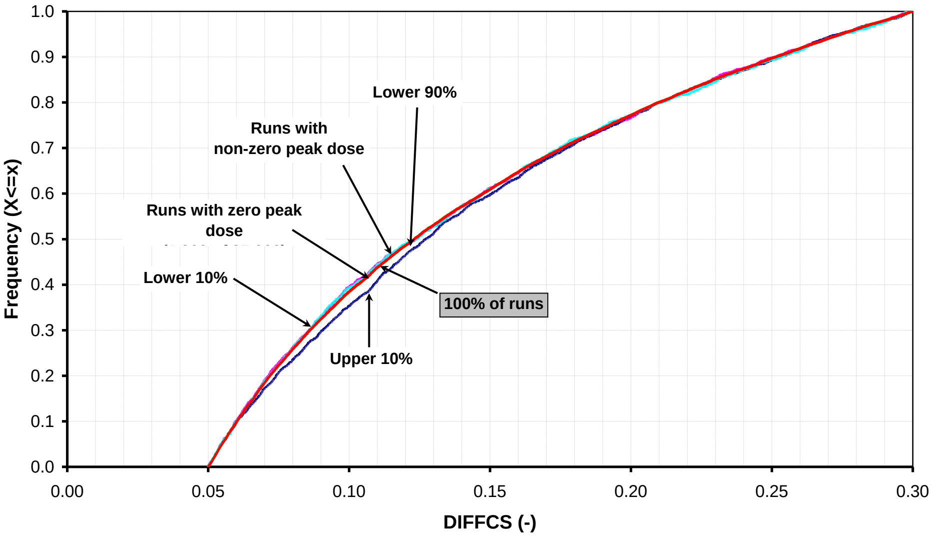


Figure 4.21 – Cumulative distribution functions of the parameter that quantifies the uncertainty in Cesium diffusion coefficient (DIFFCS) for different subsets of the 25,000 run calculation defined on the basis of peak C14 dose.

4.3.2 Numerical statistics and tests

As can be seen in Figure 4.17 to Figure 4.21 for all the parameters, even those non important, there will be some differences between the different subsets of the sample. It is necessary to define some criteria to decide when these differences are due to a real influence of the parameter on the output and when it is an statistical fluctuation. In this report, this has been done using the Smirnov, Mann-Whitney and t two-sample tests to present the values of the Smirnov, Mann-Whitney and t tests for different radionuclides and subset generation criteria.

<u>Mann-Whitney (M-W) statistic has been modified</u> using a linear transformation to make it independent of the number of runs and to take values between -1 and 1. For a calculation with n runs, two subsamples are generated: S1 with h runs and S2 with n-h runs. The original Mann-Whitney statistic is the sum of the ranks associated to the h values of the input parameter in subsample S1. The mean rank of the parameter in S1 subsample minus the mean rank of the parameter in the calculation (S1+S2) is:

$$\frac{M-W}{h} - \frac{n+1}{2}$$

which can take values between -(n-h)/2 and (n-h)/2. The normalized Mann-Whitney* statistics is defined as

$$M-W^* = \frac{\frac{M-W}{h} - \frac{n+1}{2}}{\frac{n-h}{2}}$$

A negative value of M-W* means that the runs in sample S1 (10% highest doses, for instance) are correlated with the lower values of the input parameter. A positive value of M-W* means that the runs in subsample S1 are correlated with the higher values of the input parameter. The higher the absolute value of M-W*, the stronger will be the correlation between the subsample S1 and the input parameter.

t statistic has been modified to make it independent of the number of runs (t*).

$$t^* = \frac{t}{\sqrt{n}}$$

4.3.2.1 Effect of the number of runs on the statistics

Table 4.14 and Table 4.15 present the values of the Smirnov, Mann-Whitney* and t* tests (both the p-value and the statistic itself) for the peak dose due to I129 (output) and all the random and derived parameters related with iodine (inputs) in the 10%-90% case. Results for two different calculations are shown: 5000 and 25000 runs.

With the original definitions, M-W and t statistics would change with the number of runs, but M-W* and t* do not. This allows to compare the results obtained with different numbers of runs.

The threshold for the p-value and the statistics are calculated on the basis of the p-values and the statistic for the 115 random input parameters in the model that it are known for sure to be irrelevant for the dose due to I129. The threshold for the p-value is the minimum of the p-values for the irrelevant parameters, and the threshold for a statistic is the maximum of the absolute values of the statistic for the irrelevant parameters. Both threshold are included in the last rows of and Table 4.15, where only parameters with a p-value or statistic above the applicable thresholds are included.

It can be seen that the statistics are little affected by the number of runs while a factor 5 of increase in the number of runs leads to many orders of magnitude of change in the p-values. As expected, the capability of the tests to identify small differences between the 2 subsets increases with the number of runs. Going from 5,000 to 25,000 runs decreases the threshold of the statistics by a factor slightly higher than 2, which points to a behaviour that follows the “rule of thumb”:

$$\text{threshold}\,\alpha \frac{1}{\sqrt{n}}$$

It has been found that the statistics are robust parameters that tend to the “real” value when the number of runs increases. By the contrary, the p-values of the relevant parameters tend to zero when the number of runs increases. In addition the statistics allow making a ranking of importance of the input parameters. For the reasons presented, in the next sections the results of the Monte Carlo filtering analyses are presented only in terms of the statistics for the 25,000 run calculation.

The three tests used allow identifying all the important parameters with just 5,000 runs, and the capability to identify less relevant parameters improves with 25,000 runs. The relevant parameters identified are in agreement with those obtained in the correlation and regression analyses.

Table 4.14 – Monte Carlo filtering for I129 peak doses (10%-90%).

	Smirnov				Mann-Whitney *			
	5000		25000		5000		25000	
	p-value	statistic	p-value	statistic	p-value	statistic	p-value	statistic
Waste(3)								
R1000	2E-04	0.100	3E-14	0.084	1E-04	0.104	0	0.104
GAPX	7E-84	0.459	0	0.429	0	0.564	0	0.548
Canisters (3)								
MEAN	1E-3	0.091	1E-21	0.104	1E-04	-0.106	4E-25	-0.126
SLOPE			7E-08	0.062	9E-04	0.091	2E-08	0.068
Near field (6)								
WATER NF	5E-80	0.449	0	0.428	0	0.540	0	0.530
KDBEX	1E-15	0.196	1E-61	0.177	3E-19	-0.244	2E-76	-0.225
TT NF	1E-107	0.521	0	0.504	2E-110	-0.608	0	-0.594
1/TT NF	1E-107	0.521	0	0.504	0	0.608	0	0.594
Water flow (4)								
WATER TT0	4E-94	0.487	0	0.470	1E-91	0.564	0	-0.542
KINEM POR	9E-21	0.227	2E-124	0.251	6E-29	-0.304	1E-166	-0.335
PECLET	5E-04	0.095	2E-19	0.098	3E-04	0.099	0	0.124
Granite matrix (6)								
FWA			1E-06	0.056			2E-7	-0.063
TT FF	3E-145	0.605	0	0.577	9E-145	-0.697	0	-0.690
1/TT FF	3E-145	0.605	0	0.577	0	0.697	0	0.690
DecayT	4E-145	0.605	0	0.577	0	0.697	0	0.690
TT NF+TT FF	2E-144	0.786	0	0.763	2E-235	-0.892	0	-0.881
1/(TT NF+TT FF)	2E-144	0.786	0	0.763	0	0.892	0	0.881
GAP/(TT NF+TT FF)	4E-302	0.873	0	0.865	0	0.955	0	0.955
DecayT/(TT NF+TT FF)	2E-144	0.786	0	0.763	0	0.892	0	0.881
Threshold p-value	3E-03	0.084	2E-03	0.039	2E-03	0.086	6E-03	0.033

Table 4.15 – Monte Carlo filtering for I129 peak doses (10%-90%).

	t *			
	5000		25000	
	p-value	statistic	p-value	statistic
Waste(3)				
R1000	5E-05	0.057	0	0.057

GAPX	0	0.307	0	0.297
Canisters (3)				
MEAN	1E-04	0.055	0	0.066
SLOPE	8E-04	0.047	2E-08	0.035
Near field (6)				
WATER NF	0	0.266	0	0.261
KDBEX	0	0.128	0	0.118
TT NF	0	0.211	0	0.208
1/TT NF	0	0.177	0	0.175
Water flow (4)				
WATER TT0	0	0.318	0	0.308
KINEM POR	0	0.137	0	0.149
PECLET	2E-04	0.053	0	0.067
Granite matrix (6)				
FWA	3E-03	0.042	3E-09	0.038
TT FF	0	0.276	0	0.280
1/TT FF	0	0.285	0	0.296
DecayT	0	0.277	0	0.281
TT NF+TT FF	0	0.353	0	0.352
1/(TT NF+TT FF)	0	0.657	0	0.635
GAP/(TT NF+TT FF)	0	0.903	0	0.904
DecayT/(TT NF+TT FF)	0	0.657	0	0.635
Threshold p-value	5E-03	0.040	3E-03	0.019

4.3.2.2 Results of the Monte Carlo filtering analyses

Table 4.16 to Table 4.20 show the results obtained in the Monte Carlo filtering analyses for different radionuclides and subsets generation criteria for the 25,000 runs calculation:

- Since the highest doses are of special interest in the Safety Assessment, the majority of the analyses have been done comparing the 10% of runs producing the highest doses with the remaining 90% of runs (10%-90% cases).

- For C14, I129 and Se79 the different statistics have been calculated for 2 subsets of the 10% of runs producing the lowest doses and the remainder 90% (90%-10% cases) also.

- Finally, for C14 the statistics were calculated for 2 subsets of runs: runs that cause a zero peak doses (1200 runs out of 25,000) vs. the runs that cause non zero peak doses (zero vs. non zero case).

Smirnov and t* statistics are calculated using the values of the input parameters and the output variable while M-W* statistic is calculated using ranks.

<u>10%-90% cases</u>

In the 10%-90% cases the parameters with high values of the different statistics (Smirnov. Mann-Whitney* and t*) are in good agreement with those identified as correlated with the peak doses in the correlation and regression analysis (section 4.1). The ranking of importance of the parameters is similar for the different statistics and to the ranking based on correlation coefficients, with some exceptions for the t* statistic (see 4.3.2.3).

10%-90% cases vs. 90%-10% cases

For I129, C14 and Se79 the parameters with high (absolute) values of the 3 statistics (Smirnov, Mann-Whitney* and t*) in the 10%-90% and 90%-10% cases are roughly the same. For all the parameters the Mann-Whitney* statistic has different sign in the 10%-90% and the 90%-10% cases: if the runs with highest doses are correlated with high values of a given parameters, then the runs with lowest doses are correlated with the low values of the same parameter, and vice versa. This is a logical result, consequence of the monotony of the model.

zero vs. non zeros case

C14 produces no doses in 1,200 of the 25,000 runs. For C14 the results obtained in the 90%-10% and zero vs. non zero cases are quite similar. In the disposal system in granite considered, the runs with zero dose do not correspond to a set of values of the parameters totally different from the non-zero dose runs. Zero dose runs simply are a subset of the runs with lowest peak doses.

Table 4.16 – Monte Carlo filtering for peak total dose (10% of highest doses) and I129 peak doses (10% of highest doses and 10% of lowest doses).

	Total (10%-90%)			I129 (10%-90%)			I129 (90%-10%)		
	Smir	M-W*	t*	Smir	M-W*	t*	Smir	M-W*	t*
Waste(3)									
R1000	0.081	0.101	0.055	0.084	0.104	0.057	0.256	-0.324	0.173
GAPX	0.406	0.516	0.278	0.429	0.548	0.297	0.334	-0.441	0.236
Canisters (3)									
MEAN	0.104	-0.126	0.066	0.104	-0.126	0.066			
SLOPE	0.059	0.064	0.033	0.062	0.068	0.035			
Near field (6)									
POR ANIONS BUFFER									
WATER NF	0.438	0.541	0.270	0.428	0.530	0.261	0.578	-0.677	0.208
SOLU			0.028						
KDBEI	0.162	-0.201	0.105	0.177	-0.225	0.118	0.294	0.381	0.202
TT NF	0.505	-0.593	0.208	0.504	-0.594	0.208	0.633	0.737	0.713
1/TT NF	0.505	-0.593	0.172	0.504	0.594	0.175	0.633	-0.737	0.118
Water flow (4)									
WATER TT0	0.451	-0.554	0.317	0.470	-0.542	0.308	0.161	0.212	0.108
KINEM POR	0.233	-0.310	0.136	0.251	-0.335	0.149	0.216	0.283	0.187
PECLET	0.086	0.108	0.059	0.098	0.124	0.067	0.112	-0.135	0.066
Granite matrix (6)									
FWA	0.073	-0.095	0.048	0.056	-0.063	0.038		0.035	0.020
THICKM		-0.050	0.026						
TT FF	0.574	-0.689	0.278	0.577	-0.690	0.280	0.272	0.333	0.259
1/TT FF	0.574	0.689	0.304	0.577	0.690	0.296	0.272	-0.333	0.067
DecayT	0.574	0.689	0.279	0.577	0.690	0.281	0.271	-0.332	0.259
TT NF+TT FF	0.759	-0.877	0.350	0.763	-0.881	0.352	0.733	0.882	0.743

1/(TT NF+TT FF)	0.759	0.877	0.644	0.763	0.881	0.635	0.733	-0.882	0.200
GAP/(TT NF+TT FF)	0.850	0.942	0.895	0.865	0.955	0.904	0.880	-0.950	0.200
DecayT/(TT NF+TT FF)	0.759	0.877	0.644	0.763	0.881	0.635	0.732	-0.881	0.200
Threshold statistic	0.039	0.034	0.020	0.039	0.033	0.019	0.033	0.034	0.018

Table 4.17 – Monte Carlo filtering for Cl36 and Cs135 peak doses (10%-90%).

	Cl36			Cs135		
	Smir	M-W*	t*	Smir	M-W*	t*
Waste(3)						
R1000	0.096	0.122	0.065	0.132	0.170	0.089
GAPX	0.481	0.596	0.326		0.038	0.020
Canisters (3)						
MEAN	0.143	-0.177	0.092			
SLOPE	0.090	0.109	0.057			
Near field (6)						
POR ANIONS BUFFER	0.054	-0.067	0.035			
WATER NF	0.177	0.212	0.072	0.323	0.393	0.164
KDBEX				0.082	-0.111	0.059
TT NF	0.187	-0.219	0.128	0.332	-0.412	0.180
1/TT NF	0.187	0.219	0.063	0.332	0.412	0.146
Water flow (4)						
WATER TT0	0.629	-0.721	0.434	0.487	-0.582	0.335
KINEM POR	0.237	-0.331	0.150			
PECLET	0.079	0.104	0.056			
Granite matrix (6)						
FWA	0.056	-0.062	0.032	0.381	-0.507	0.194
THICKM	0.042	-0.038	0.022	0.284	-0.361	0.174
MATCOR				0.301	-0.374	0.181
KDRX				0.196	-0.268	0.126
TT FF	0.727	-0.828	0.318	0.874	-0.938	0.139
1/TT FF	0.727	0.828	0.472	0.874	0.938	0.488
DecayT	0.727	0.828	0.387	0.874	0.938	1.000
TT NF+TT FF	0.755	-0.861	0.330	0.916	-0.986	0.147
1/(TT NF+TT FF)	0.755	0.861	0.562	0.916	0.986	0.986
GAP/(TT NF+TT FF)	0,852	0.939	0.811	0.864	0.967	0.915
DecayT/(TT NF+TT FF)	0.753	0.860	0.802	0.913	0.984	0.742
Threshold statistic	0.041	0.035	0.019	0.034	0.034	0.018

Table 4.18 – Monte Carlo fIltering for C14 peak doses.

	10%-90%			zeros vs. non zeros			90%-10%		
	Smir	M-W*	t*	Smir	M-W*	t*	Smir	M-W*	t*
Waste(3)									
METAL LIFETIME	0.042	-0.048	0.029						
GAPC						0.014			
Canisters (3)									

FAIL	0.039	0.040							
MEAN	0.051	-0.054	0.028	0.052	0.045	0.016			
Near field (6)									
WATER NF	0.187	0.237	0.108		-0.038	0.013	0.046	-0.046	0.018
KDBEC	0.128	-0.171	0.089	0.045		0.014	0.039	0.034	
TT NF	0.221	-0.297	0.125		0.048	0.016	0.045	0.051	0.030
1/TT NF	0.221	0.297	0.079		-0.048		0.045	-0.051	
Water flow (4)									
WATER TT0	0.748	-0.842	0.535	0.311	0.432	0.165	0.313	0.433	0.240
KINEM POR	0.157	-0.218	0.107	0.061	0.067	0.030	0.083	0.096	0.054
PECLET	0.121	-0.158	0.074	0.346	0.448	0.170	0.314	0.413	0.220
Granite matrix (6)									
FWA	0.229	-0.310	0.136	0.542	0.685	0.329	0.475	0.619	0.384
THICKM	0.184	-0.237	0.117	0.392	0.492	0.195	0.356	0.428	0.234
MATCOR	0.188	-0.237	0.118	0.366	0.476	0.187	0.317	0.401	0.219
KDRC	0.119	-0.160	0.082	0.259	0.352	0.133	0.216	0.299	0.159
TT FF	0.873	-0.961	0.174	0.915	0.969	0.850	0.866	0.950	0.805
1/TT FF	0.873	0.961	0.949	0.915	-0.969	0.112	0.866	-0.950	0.164
DecayT	0.873	0.961	0.602	0.915	0.969	0.042	0.866	-0.950	0.062
TT NF+TT FF	0.649	-0.796	0.215	0.813	0.923	0.707	0.703	0.853	0.686
1/(TT NF+TT FF)	0.649	0.796	0.755	0.813	-0.923	0.132	0.703	-0.853	0.188
DecayT/(TT NF+TT FF)	0.888	0.977	0.287	0.916	-0.969	0.021	0.867	-0.950	0.031
Threshold statistic	0.038	0.033	0.023	0.045	0.038	0.012	0.037	0.031	0.016

Table 4.19 – Monte Carlo filtering for Se79 peak doses.

	10%-90%			90%-10%		
	Smir	M-W*	t*	Smir	M-W*	t*
Near field (6)						
WATER NF	0.586	0.707	0.420	0.414	-0.520	0.181
SOLSE	0.625	0.753	0.488	0.415	-0.525	0.192
TT NF	0.501	-0.621	0.208	0.347	0.454	0.259
1/TT NF	0.501	0.621	0.157	0.347	-0.454	0.076
MF NF	0.910	0.961	0.967	0.528	-0.679	0.133
Water flow (4)						
WATER TT0	0.139	-0.169	0.092	0.214	0.287	0.156
KINEM POR	0.046	-0.047	0.028	0.044	0.047	0.022
Granite matrix (6)						
FWA	0.130	-0.177	0.087	0.308	0.414	0.244
THICKM	0.093	-0.118	0.064	0.252	0.320	0.178
MATCOR	0.087	-0.111	0.058	0.220	0.284	0.154
KDRSE	0.098	-0.122	0.065	0.242	0.332	0.190
TT FF	0.272	-0.355	0.101	0.544	0.683	0.549
1/TT FF	0.272	0.355	0.088	0.544	-0.683	0.119
DecayT	0.272	0.355	0.190	0.544	-0.683	0.474
TT NF+TT FF	0.392	-0.529	0.116	0.597	0.776	0.575
1/(TT NF+TT FF)	0.392	0.529	0.243	0.597	-0.776	0.166

DecayT/(TT NF+TT FF)	0.367	0.501	0.234	0.567	-0.752	0.154
MF NF*DecayT	0.968	0.996	1.212	0.974	-0.995	0.139
Threshold statistic	0.037	0.034	0.018	0.036	0.029	0.016

Table 4.20 – Monte Carlo filtering for 4n+1 and 4n+2 series peak doses (10%-90%).

	4n+1 series			4n+2 series		
	Smir	M-W*	t*	Smir	M-W*	t*
Near field (6)						
WATER NF	0.220	0.280	0.127	0.272	0.344	0.162
SOLNP	0.227	0.304	0.125			
SOLU				0.363	0.476	0.183
KDBEBP			0.022			
DIFFNP		0.046	0.023			
TT NF	0.214	-0.278	0.114	0.262	-0.328	0.134
1/TT NF	0.214	0.278	0.114	0.262	0.328	0.130
MF NF	0.311	0.415	0.120	0.492	0.602	0.170
Water flow (4)						
WATER TT0	0.490	-0.577	0.332	0.433	-0.522	0.298
KINEM POR						
PECLET	0.106	-0.135	0.063	0.098	-0.121	0.056
Granite matrix (6)						
FWA	0.395	-0.530	0.200	0.346	-0.463	0.183
THICKM	0.290	-0.369	0.178	0.253	-0.324	0.156
MATCOR	0.301	-0.378	0.182	0.271	-0.344	0.165
KDRNP	0.169	-0.221	0.104			
KDRRA				0.064	-0.074	0.037
KDRTH	0.090	-0.107	0.047	0.060	-0.069	0.029
KDRU				0.123	-0.149	0.075
TT FF	0.825	-0.931	0.134	0.765	-0.865	0.147
1/TT FF	0.825	0.931	0.542	0.765	0.865	0.315
DecayT	0.825	0.931	0.692			
TT NF+TT FF	0.570	-0.736	0.162	0.666	-0.823	0.167
1/(TT NF+TT FF)	0.570	0.736	0.665	0.666	0.823	0.677
Decay/(TT NF+TT FF)	0.842	0.950	0.449			
MF NF*DecayT	0.883	0.976	0.288			
Threshold statistic	0.040	0.037	0.021	0.033	0.033	0.017

For radionuclides that have runs with zero peak dose, similar results are obtained using two different sets of subsets: zero vs. non zero runs and 10% highest doses vs. 90% lowest doses, but in general the comparison of the 10%-90% subsets is better to identify less relevant parameters. In the repository in granite considered it is appropriate to use the 10% highest runs and the 90% lowest runs as sub-samples to test non parametric statistics.

4.3.2.3 *Anomalies of the t* statistic*

In some cases the t* statistic presents a strange behaviour, as it can be observed in this summary of results for Se79

Table 4.21 – Summary of results for Se79.

	CC SRC	RCC SRRC	10%-90%			90%-10%		
			Smir	M-W*	t*	Smir	M-W*	t*
WATER NF	0.437	0.582	0.586	0.707	0.420	0.414	-0.520	0.181
SOLSE	0.492	0.627	0.625	0.753	0.488	0.415	-0.525	0.192
MF NF	0.837	0.863	0.910	0.961	0.967	0.528	-0.679	0.133
DecayT	0.220	0.395	0.272	0.355	0.190	0.544	-0.683	0.474
MF NF*DecayT	0.974	0.995	0.968	0.996	1.212	0.974	-0.995	0.139

Since Se79 is a solubility controlled radionuclide, their peak doses are controlled by the product SOLSE·WATER NF=MF NF and this product should have more influence on the peak doses due to Se79 than any of the two factors separately. This behaviour is observed in 7 of the columns in Table 4.21, but on the last column the t* statistic of MF NF is smaller than the t* statistic for SOLSE and WATER NF.

Another anomaly is observed in Table 4.21 with MF NF·DecayT for Se79. This derived parameter has the highest value (in absolute value) in the first 7 columns, but the t* statistic for MF NF·DecayT is much smaller than for DecayT.

For C14 the t* statistic provides strange results also:

- The "derived parameter" DecayT/(TT NF+TT FF) is the parameter with highest CC and RCC for the peak dose due to C14 (Table 4.10), and
- DecayT/(TT NF+TT FF) always has the highest value of the Smirnov and M-W* statistics in the Monte Carlo filtering for C14 (Table 4.18), but
- there are several parameters with a value of the t* statistic much greater than DecayT/(TT NF+TT FF) in the Monte Carlo filtering for C14 (Table 4.18).

The t-test is based on the sampling distribution of the mean of normal variables, and none of the random input or derived parameters in the model have normal distributions. Perhaps this is the cause of the strange results of the t* statistic presented in this section. This topic will be further investigated.

On the basis of the results obtained, it is recommended to use only Smirnov and Mann-Whitney test for Monte Carlo filtering in values and ranks respectively. Although t* statistic in general produces results consistent with the other statistics, the existence of some strange results makes advisable not to use it, and perform the Monte Carlo filtering in values using only the Smirnov test.

4.4 Cobweb plots

4.4.1 Cobweb plots for I129

In this report only conditional cobweb plots have been used, where only a fraction of the runs obtained are represented. Due to the relevance of the peak doses for the Performance Assessment, analyses are focused on the runs that produce the 10% highest peak doses.

Figure 4.22 and Figure 4.23 are the conditional cobweb plots for the 20 independent random input parameters that can affect the doses due to I129. Visual inspection allows to identify GAPI and WATER NF as important parameters and KDBEI, WATER TT0 and KINEM POR as secondary parameters. These 5 parameters agree with the parameters with correlation and regression coefficients higher than 0.2 (in absolute value) for I129 peak dose (Table 4.3).

Figure 4.24 is the conditional cobweb plot for 4 “derived parameters” for iodine. It can be seen that the highest doses due to I129 are strongly correlated with the low values of the travel time through the near field (TT NF) and far field (TT FF), with the low values of the total travel time (TT NF+TT FF) and with the high values of GAPI/(TT NF+TT FF).

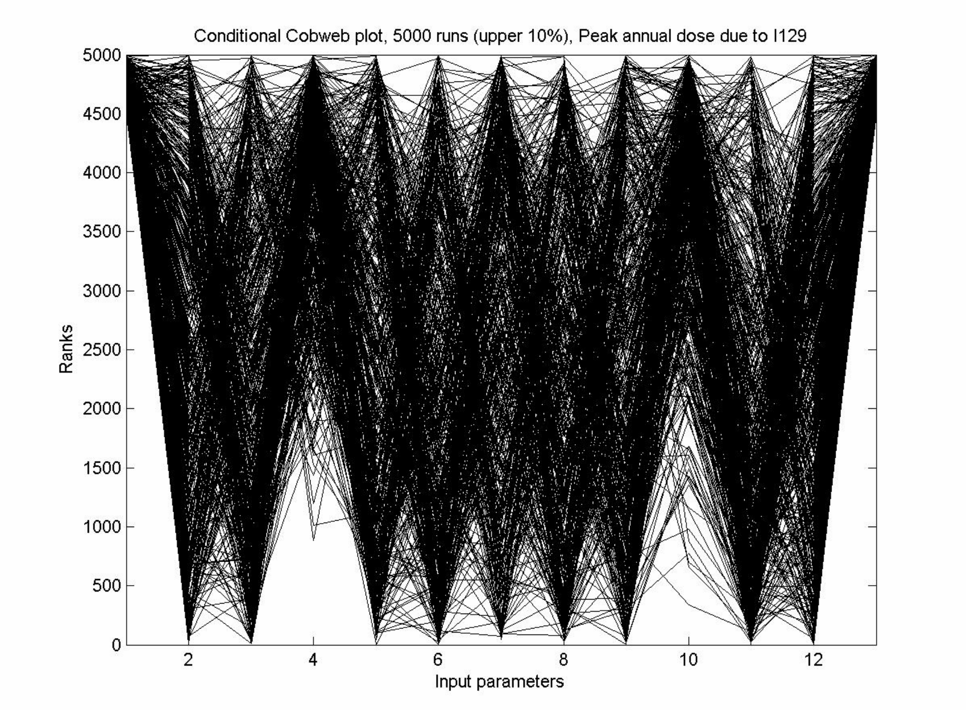


Figure 4.22 – Conditional cobweb plot (ranks, condition: 10% largest output values) for the peak dose due to I129 and the parameters of the waste, canister and near field (see Table 3.1). Parameter 4 is GAPI, parameter 10 is WATER NF and parameter 11 is KDBEI.

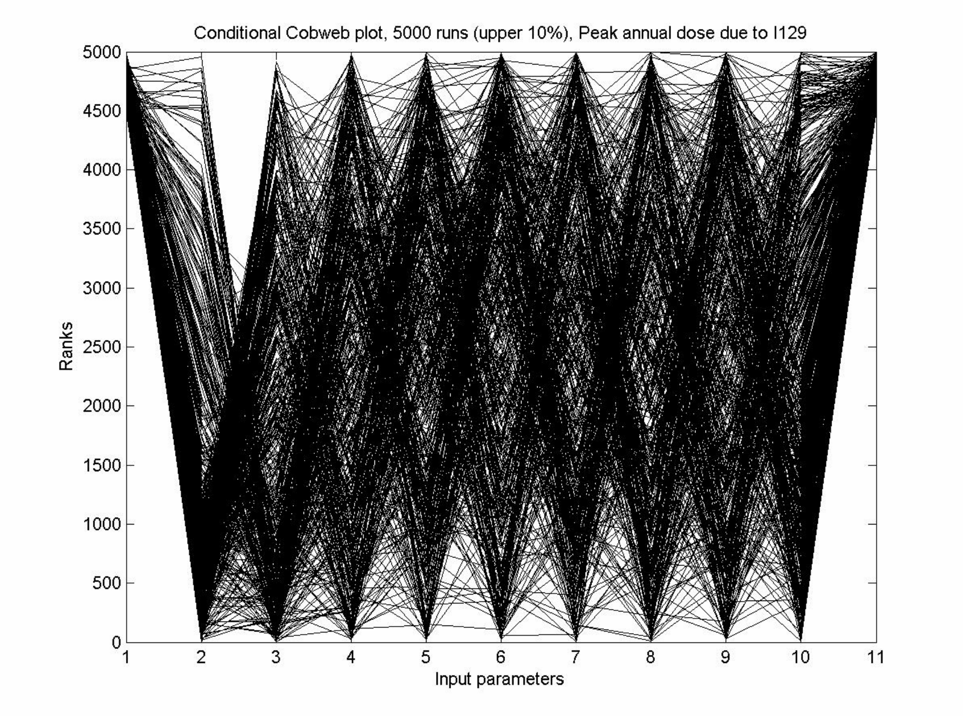


Figure 4.23 – Conditional cobweb plot (ranks, condition: 10% largest output values) for the peak dose due to I129 and the parameters of the far field (see Table 3.1). Parameter 2 is WATER TT0 and parameter 3 is KINEM POR.

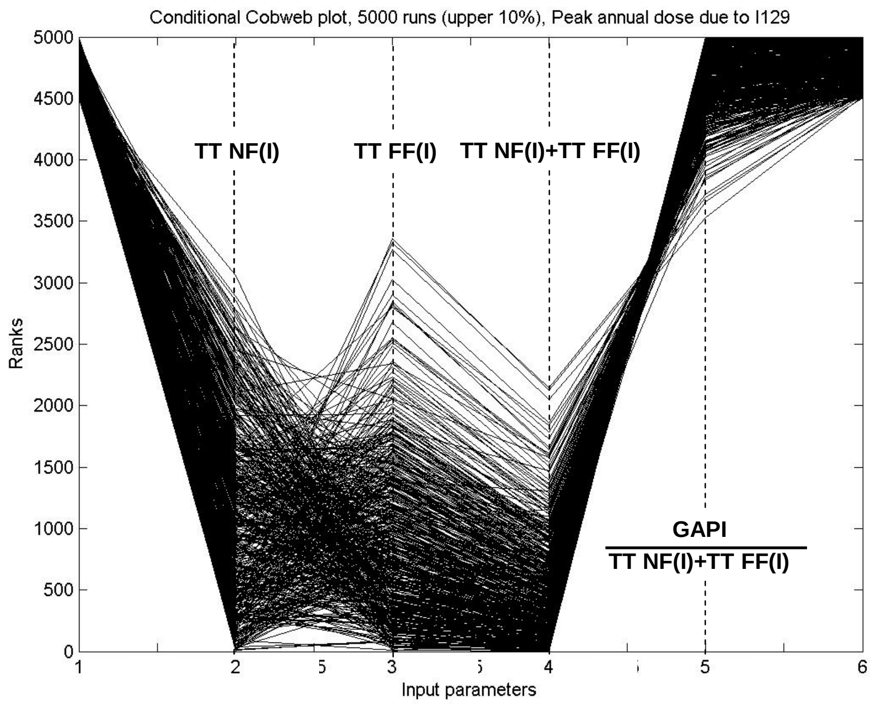


Figure 4.24 – Conditional cobweb plot (ranks, condition: 10% largest output values) for the peak dose due to I129 and the “derived parameters” for iodine (see 4.1.4).

Cobweb plots in Figure 4.22 and Figure 4.24 allow identifying the parameters strongly correlated with the peak doses due to I129, but less important parameters (such as R1000 and PECLET) can not be identified.

4.4.2 Potential visual artefacts in the cobweb plots.

In the calculations performed it was found that the cobweb plots could produce visual artefacts. Figure 4.25 is a rank based conditional cobweb plot where the output variable is the peak dose due to I129 and the conditioning event is getting one of the 10% highest values. The output parameter is represented on columns 1 and 8. Parameters represented on columns 5 and 7 are not important (DIFFAG and DIFFAM), while parameters on columns 2 and 4 are very important (WATER NF and GAPI). On columns 3 and 6 *the same* input parameter is represented (VC1). Column 6 shows this parameter as non-important, since it does not seem to be any specific relation with the 10% highest values of the output. Column 3 shows some relation between the small values of this input and the 10% highest values of the output.

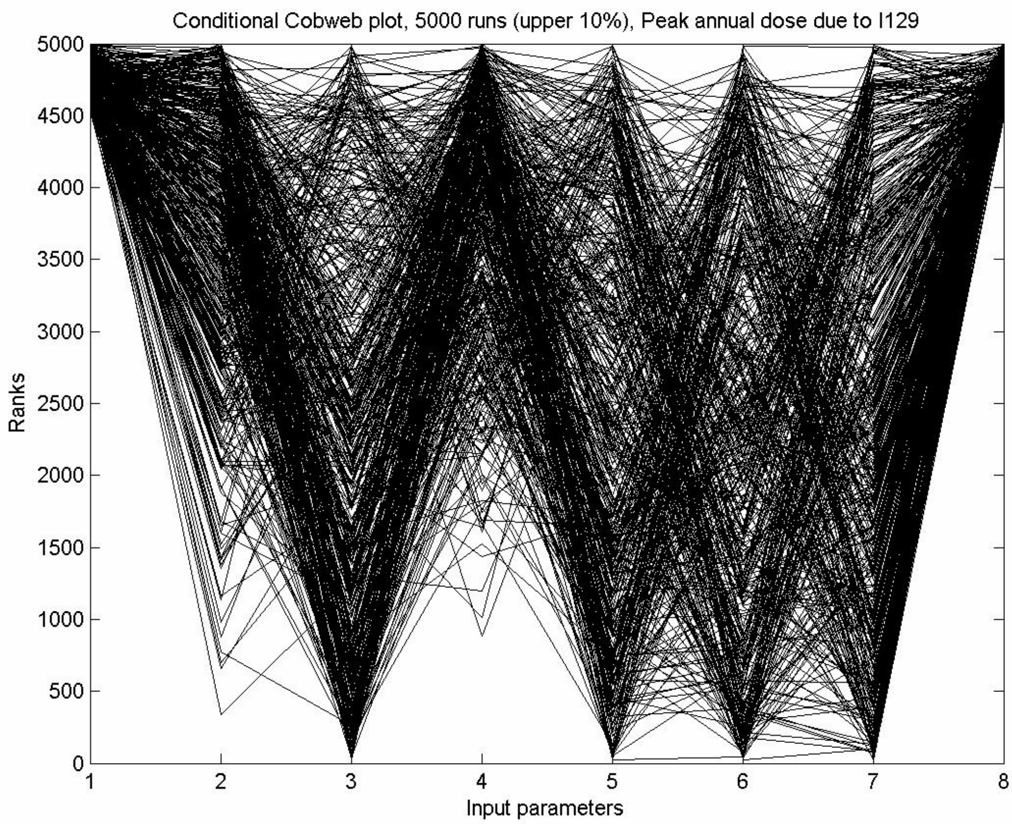


Figure 4.25 – Conditional cobweb plot (ranks, condition: 10% largest output values) for the peak dose due to I129 and a few selected parameters.

There is a visual artefact in column 3, caused by having on both sides two parameters with a strong relation (same type; larger values related to larger values) with the output. Mann-

Whitney test indicates that VC1 is not a relevant parameter, and this is the information that must be trust. This class of visual artefacts can be avoided by placing the parameter of interest between two non important parameters.

4.4.3 Alternative graphic representation

As shown in section 4.4.1, conditional cobweb plots allow identifying the parameters strongly correlated with a region of the output variable (10% of runs with highest peak doses, for instance). Cobweb plots graphically show the ranges of values of the input parameters that correspond to the selected region of the output variable: high values of parameter A, low values of parameter B,…

Due to the great number of lines represented (one per run) cobweb plots can only present a few parameters in X axis, become blurred and only strong correlations can be identified visually. In addition cobweb plots can produce visual artefacts as described in section 4.4.2.

An alternative graphical representation has been developed where only the mean ranks of the parameters in the sub-sample selected (10% of runs with highest peak doses, for instance). The mean ranks for all the parameters relevant for a radionuclide (about 20 input parameters plus 3 to 5 derived parameters) can be represented in a single figure. In addition, the results for several conditions on the output variable can be represented in the same figure: in our case the results for the 10% or runs with highest doses (Upper 10%) and the lowest doses (Lower 10%) are represented together. The resulting figures are Figure 4.26 for I129 and Figure 4.27 for Se79.

Parameters that have no effect on the highest/lowest doses have a mean rank around 12500.5. Mean ranks close to the 1250.5 or 23750.5 indicate a strong correlation of the parameter with the he higher the correlation of the parameter. On the basis of the mean ranks of the non-relevant parameters it is possible to identify the range of values of the mean rank that is not statistically significant: 12500.5±379 for I129 and 12500.5±378 for Se79. This range is represented as a grey band in Figure 4.26 and Figure 4.27, where all the parameters whose mean ranks are statistically significant for any of the two conditions are labelled.

Significant parameters in are the same that were found important in the correlation and regression analyses for the peak dose due to I129, with a similar ranking. In general, for a given parameter the points for the upper and lower 10% of runs present some symmetry (using the line Mean rank=12500.5 as symmetry axis): if the high values of the parameter are correlated with the higher/lower doses, then the low values of the parameter are correlated with the lower/higher doses. Nevertheless, the importance of a given parameter can be different for higher and lower doses: for instance, the relevance of R1000 for the highest peak doses due to I129 is much smaller than for the lowest peak doses due to I129.

Similar results are obtained for Se79 (Figure 4.27). It is noteworthy that the 2500 runs with the highest/lowest peak dose due to Se79 are roughly the same than the 2500 runs with the highest/lowest values of MF NF·DecayT. The mean rank of MF NF·DecayT is 1304 in the

10% of runs with lowest peak doses (mean rank 1250.5) and 23700 in the 10% of runs with highest doses (mean rank 23750.5).

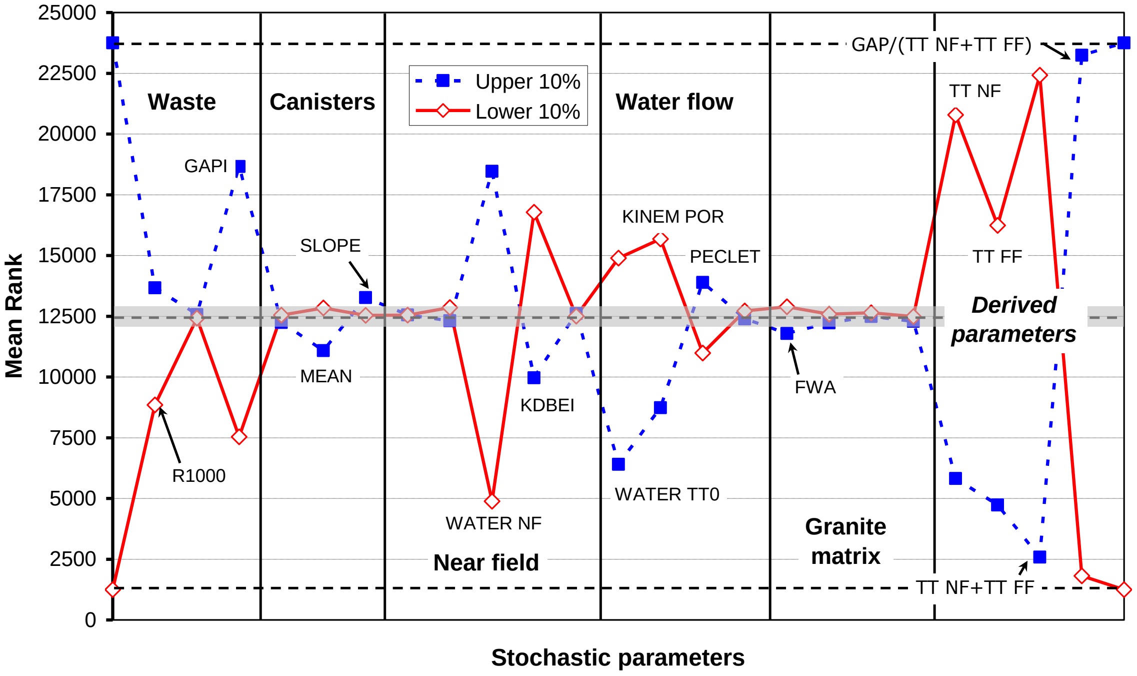


Figure 4.26 – Mean ranks for the independent and derived parameters for Iodine in the 10% of runs with the highest and lowest peak dose due to I129 (Upper and Lower 10%).

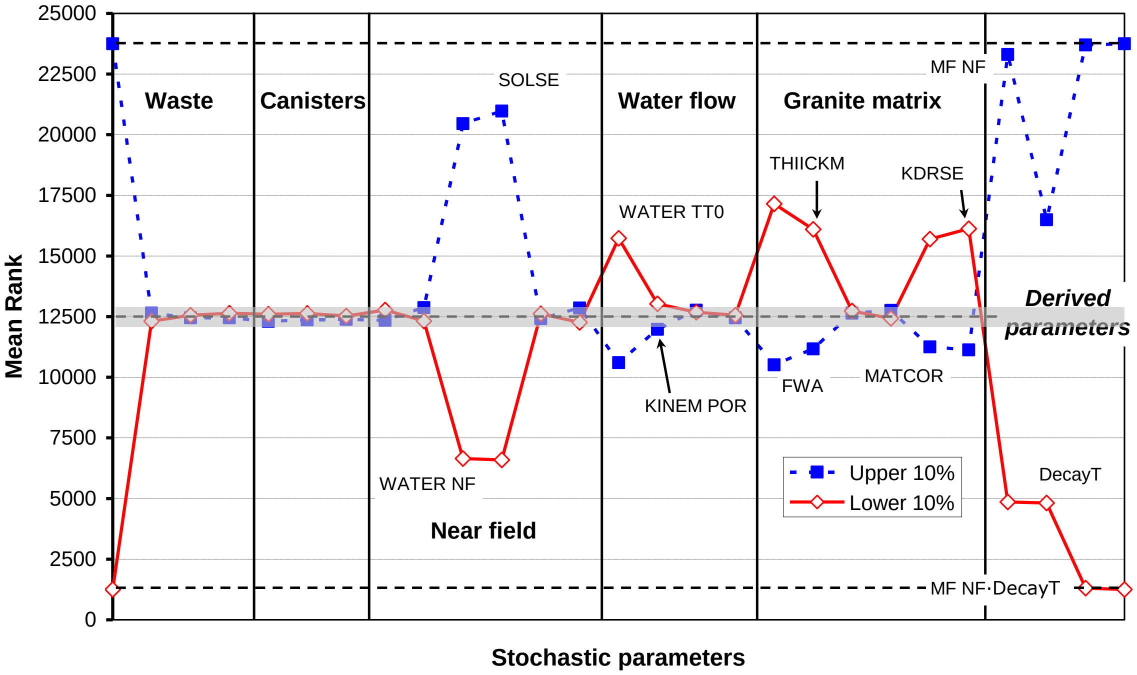


Figure 4.27 – Mean ranks for the independent and derived parameters for Selenium in the 10% of runs with the highest and lowest peak dose due to Se79 (Upper and Lower 10%).

It is possible to define a magnitude (X) to summarise the information presented in Figure 4.26 and Figure 4.27 for each parameter.

$$X = \frac{\text{Mean rank (Upper 10\%) - Mean rank (Lower10\%)}}{0.9 \cdot n}$$

where n is the number of runs in the calculation (25,000 in previous examples). Negative values of X mean that the input parameter and output variable are negatively correlated: high values of the parameter lead to low values of the output and vice versa. High absolute values of X point to a strong influence of the parameter on the output variable, while low absolute values correspond to a weak influence.

The values of magnitude X for I129 and Se79 in the 25,000 runs calculation are shown in Table 4.22. The ranking of importance of the difference parameters on the basis of the X parameters is in good agreement with the ranking obtained in section 4.6 with different statistics.

Table 4.22 – Value of magnitude X for the peak doses due to I129 and Se79.

	I129	Se79
Waste(3)		
R1000	0.214	
GAPI	0.495	
Canisters (3)		
MEAN	-0.078	
SLOPE	0.032	
Near field (6)		
WATER NF	0.604	0.614
SOLSE		0.639
KDBEX	-0.303	
TT NF	-0.665	
MF NF		0.820
Water flow (4)		
WATER TT0	-0.377	-0.228
KINEM POR	-0.309	-0.047
PECLET	0.129	
Granite matrix (6)		
FWA	-0.049	-0.295
THICKM	-0.033	-0.219
MATCOR		-0.198
KDRSE		-0.222
TT FF	-0.511	
DecayT		0.519
TT NF+TT FF	-0.881	
GAP/(TT NF+TT FF)	0.952	
MF NF·DecayT		0.995

4.5 Contribution to the sample mean (CSM) plots.

In the CSM plots the y-axis represents the fraction of the output variable (in our calculations, peak dose due to a radionuclide) mean corresponding to values of an input parameter smaller or equal to a given quantile while the x-axis represents the empirical cumulative distribution function of the parameters.

Figure 4.28 is an example of CSM plot. **A** point means that the 34% highest values of WATER TT0 are responsible for only 10% of the mean of the peak dose due to Cs135. **B** point means that the 10% lowest values of WATER TT0 are responsible for 37% of the mean of the peak dose due to Cs135. If the curve of a parameter is always close to the diagonal, it means that no particular region of the input parameter controls the mean of the output variable.

The test statistic adopted is the "maximum vertical distance to the diagonal" (D_m) as shown in Figure 4.28. Low values of D_m mean that all the regions of the input variable have a similar contribution to the mean of the output, while high values of D_m mean that the mean of the output is controlled by a small region of the input variable.

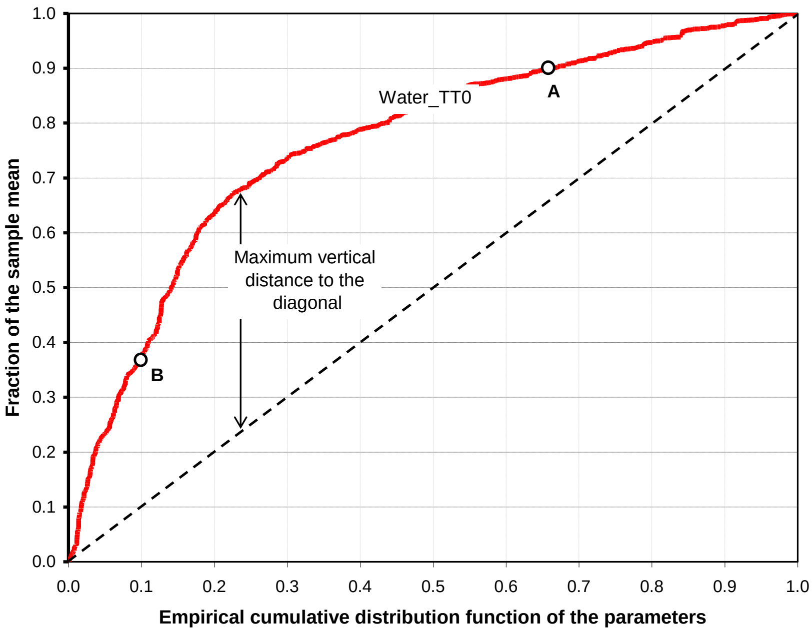


Figure 4.28 – CSM plot for input variable WATER TT0 and output variable "peak dose due to Cs135".

The test statistic (D_m) has been calculated for the peak doses of all the radionuclides identified in Figure 4.1 as producing mean doses higher than 10^{-11} Sv/yr and all the input parameters of the model, including the "derived parameters" identified in section 4.1.4.

For each radionuclide, the "threshold" value above which the calculated value of D_m is statistically significant is obtained in two different ways: the "predicted threshold" is obtained following the method proposed by R. Bolado [4] and the "empirical threshold" is the maximum value of the test statistic (D_m) for the (more than 100) random input parameters that are known for sure to have no influence on the peak dose due to a given radionuclide. There is a good agreement between the “predicted” and “empirical” threshold, as can be seen in Table 4.23 to Table 4.26.

Table 4.23 to Table 4.26 show the values of the test statistic for all the parameters with D_m higher than the “threshold” value in the 5000 runs calculation with Monte Carlo sampling (5000MC). CSM plots for several radionuclides with different behaviours are presented in Figure 4.29 to Figure 4.34 also. Due to the similarities between radionuclides, some CSM plots are not included: plots for Cl36 and the total dose are very similar to the I129 plot, Pd107 plot is similar to Se79 plot and 4n and 4n+3 plots are similar to 4n+2 plot.

Low values of the test statistic mean that the whole range of the input parameter has a similar contribution to the mean of the peak dose. By the contrary, high values of the statistic mean that the contribution of the parameter to the mean is controlled by a small range of values.

For I129 the test statistic takes small values for all the input parameters. As expected, Figure 4.29 shows that the mean of the peak dose due to I129 is not controlled by a narrow region of values of any parameter.

By the contrary, for Ni59 the mean of the peak dose is controlled by the 5% of runs with lowest values of the travel time in the far field (TT FF), as can be seen in Figure 4.31 and the test statistic value of 0.946 indicates.

The importance of TT FF for Ni59 was already identified in the correlation analysis and the Monte Carlo filtering analysis but the CSM plot allows to identify exactly the range of values of TT FF that really control the doses. This class of information can be useful to guide R&D efforts: if the lowest 5% of values of TT FF for Nickel can be ruled out (by additional investigations) the doses due to Ni59 will decrease several orders of magnitude.

Table 4.23 – Test statistic for CSM of the peak total dose and the peak dose due to I129, Cl36 and Cs135 for different parameters.

5000 runs	TOTAL	I129	Cl36	Cs135
Waste (3)				
R1000	0.042	0.044	0.029	0.146
GAPX	0.121	0.125	0.125	0.097
Canisters (3)				
MEAN	--	0.018	0.032	--
SLOPE	0.019	--	0.017	--
Near field (6)				
POR ANIONS BUFFER	--	--	0.019	--
WATER NF	0.163	0.156	0.045	0.311
SOLU	0.018	--	--	--
KDBE	0.073 (I)	0.075	--	0.115
TT NF (1/TT NF)	0.181 (I)	0.176	0.048	0.309
Water flow (4)				
WATER TT0	0.129	0.118	0.181	0.447
KINEM POR	0.084	0.088	0.125	--

PECLET	0.035	0.036	0.031	--
Granite matrix (6)				
FWA	0.025	0.018	0.018	0.363
THICKM	--	--	0.017	0.260
MATCOR	--	--	--	0.290
KDR	--	--	--	0.186
TT FF (1/TT FF)	0.161 (I)	0.157	0.233	0.711
TT NF+TT FF	0.240 (I)	0.235	0.237	0.722
GAP/(TT NF+TT FF)	0.262 (I)	0.257	0.261	0.700
Decay/(TT NF+TT FF)	0.240 (I)	0.234	0.237	0.725
Predicted threshold	0.018	0.016	0.016	0.078
Empirical threshold	0.018	0.017	0.016	0.077

Table 4.24 – Test statistic for CSM of the peak dose due to C14, Ni59, Sn126 and Tc99 for different parameters.

5000 runs	C14	Ni59	Sn126	Tc99
Waste (3)				
R1000	--		(0.166)	--
Canisters (3)				
FAIL	(0.192)		--	--
Near field (6)				
WATER NF	0.340		0.469	0.413
SOLX				0.258
KDBEX	0.379		0.441	--
TT NF (1/TT NF)	0.484	--	0.577	0.427
MF NF	--	--	0.376	0.451
Water flow (4)				
WATER TT0	0.773	0.806	0.539	0.602
KINEM POR	0.270	--	--	--
Granite matrix (6)				
FWA	0.341	0.607	0.285	0.406
THICKM	(0.185)	0.492	0.213	0.353
MATCOR	0.262	0.430	0.272	0.295
KDRX	--	--	0.341	0.353
TT FF (1/TT FF)	0.857	0.946	0.749	0.843
TT NF+TT FF	0.820	0.836	0.686	0.744
DecayT/(TT NF+TT FF)	0.874	0.949	0.825	0.859
MF NF*DecayT	--	0.952	0.782	0.867
Predicted threshold	0.177	0.281	0.145	0.159
Empirical threshold	0.248	0.307	0.173	0.176

Table 4.25 – Test statistic for CSM of the peak dose due to Se79 and Pd107 for different parameters.

5000 runs	Se79	Pd107
Waste (3)		
GAP	0.058	--
Canisters (3)		
Near field (6)		
POR ANIONS BUFFER	0.086	--
WATER NF	0.420	0.480
SOLX	0.454	0.583
TT NF (1/TT NF)	0.380	0.396
MF NF	0.581	0.679
Water flow (4)		
WATER TT0	0.116	0.149
Granite matrix (6)		
FWA	0.131	0.114
THICKM	0.105	0.119
MATCOR	0.092	0.111
KDRX	0.071	0.104
TT FF (1/TT FF)	0.220	0.246
TT NF+TT FF	0.302	0.415
DecayT/(TT NF+TT FF)	0.285	0.398
MF NF*DecayT	0.615	0.708
Predicted threshold	0.054	0.087
Empirical threshold	0.049	0.085

Table 4.26 – Test statistic for CSM of the peak dose due to the 4 actinide series for different parameters.

5000 runs	4n series (U236)	4n+1 series (Np237)	4n+2 series (U238)	4n+3 series (U235)
Waste (3)				
Canisters (3)				
Near field (6)				
WATER NF	0.463	0.437	0.469	0.507
SOLCM	0.185	--	0.203	--
SOLNP		0.460		
SOLU	0.554		0.618	0.452
MF NF	0.658	0.580	0.714	0.636
Water flow (4)				
WATER TT0	0.470	0.577	0.497	0.560
Granite matrix (6)				
FWA	0.413	0.442	0.429	0.484
THICKM	0.242	0.366	0.269	0.330
MATCOR	0.342	0.351	0.340	0.367
KDRPA	--	--	--	0.383
KDRTH	--	0.325	--	--
KDRU	0.177	--	--	--
TT FF (1/TT FF)	0.768	0.820	0.769	0.790

TT NF+TT FF	0.777	0.689	0.787	0.786
DecayT/(TT NF+TT FF)	--	0.833	--	--
MF NF*DecayT	--	0.866	--	--
Predicted threshold	0.165	0.207	0.197	0.277
Empirical threshold	0.153	0.218	0.173	0.301

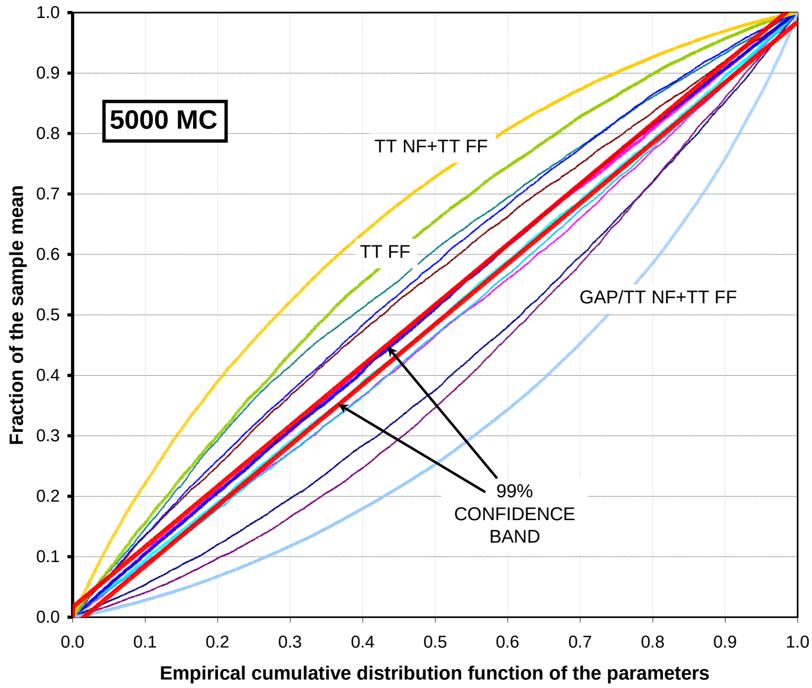


Figure 4.29 – CSM plot for relevant parameters and the peak dose due to I129 (output variable).

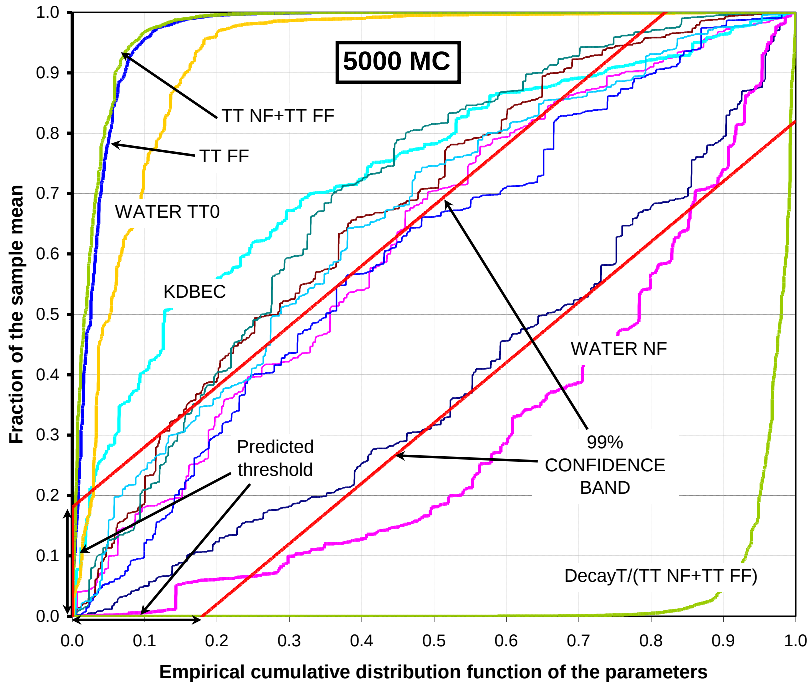


Figure 4.30 – CSM plot for relevant parameters and the peak dose due to C14 (output variable).

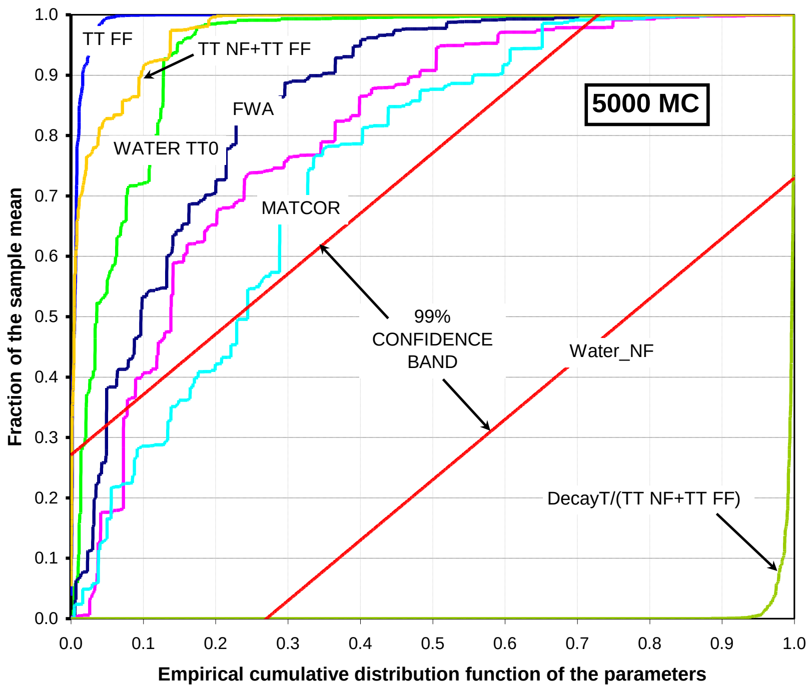


Figure 4.31 – CSM plot for relevant parameters and the peak dose due to Ni59 (output variable).

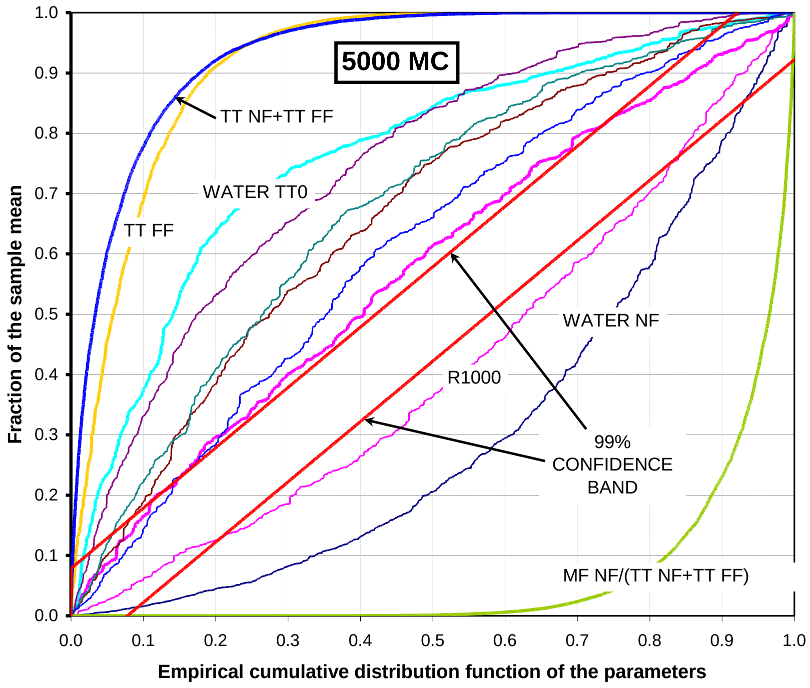


Figure 4.32 – CSM plot for relevant parameters and the peak dose due to Cs135 (output variable).

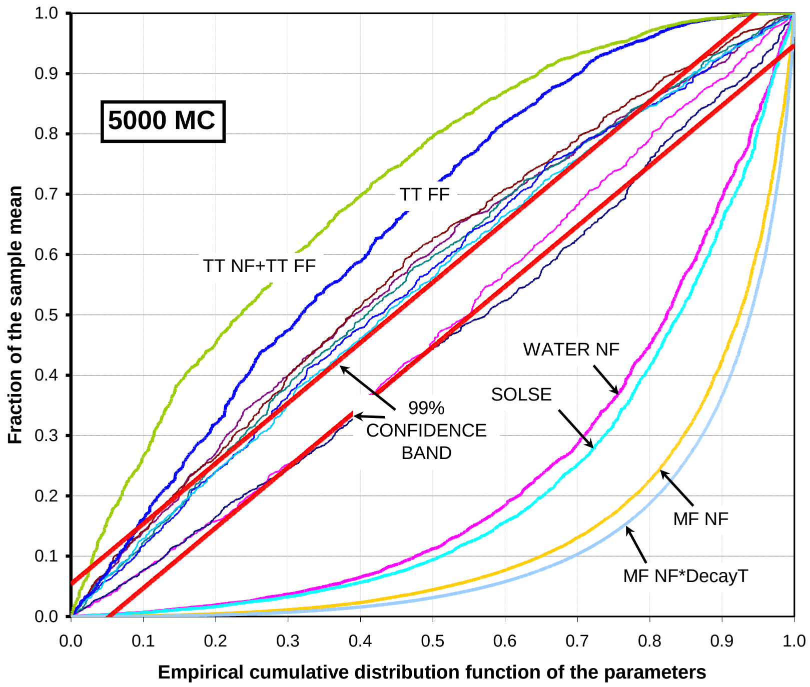


Figure 4.33 – CSM plot for relevant parameters and the peak dose due to Se79 (output variable).

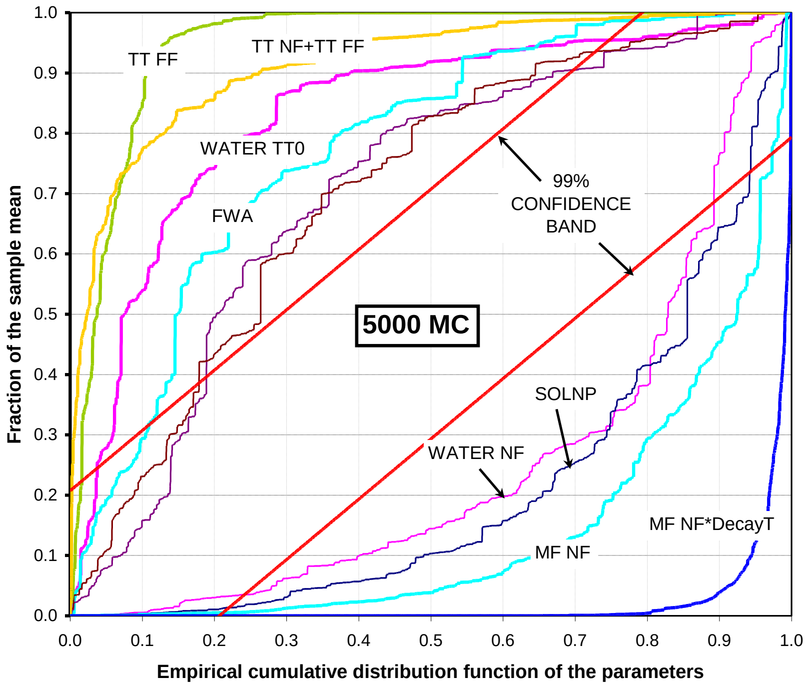


Figure 4.34 – CSM plot for relevant parameters and the peak dose due to 4n+1 series (output variable).

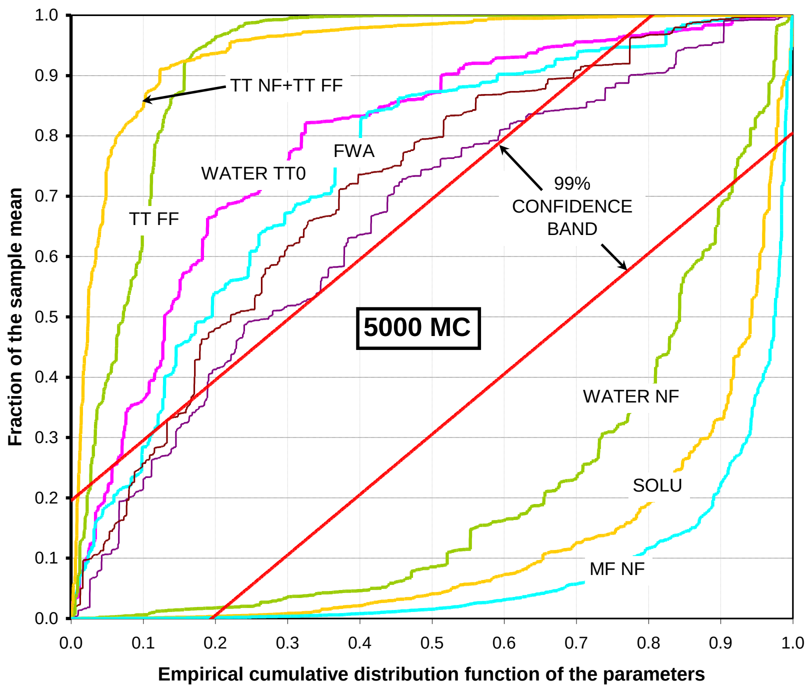


Figure 4.35 – CSM plot for relevant parameters and the peak dose due to 4n+2 series (output variable).

For some radionuclides the confidence bands obtained and shown in the figures are very wide in the 5000 runs calculations. The width probably will decrease if the number of runs increases, but when trying to do the analysis with the 25,000 runs calculation MATLAB crashed due to a memory problem during the calculation of the “predicted threshold”. As a consequence, in the 25,000 runs calculations only the “empirical thresholds” are calculated.

Table 4.27 to Table 4.29 show the values of the test statistic for CSM in the 25,000 runs calculation for the radionuclides with high “thresholds” in the 5,000 runs. In general, a factor 2 or 3 of reduction of the “empirical threshold” is observed, consistent with the rule of thumb:

$$\text{threshold}\,\alpha\,\frac{1}{\sqrt{n}}$$

For Ni53 the reduction in the “empirical threshold” when going from 5,000 to 25,000 runs is much smaller: from 0.307 to 0.253. The reason of this result is that the mean dose due to Ni59 is controlled by just a few runs. In the 5,000 runs calculation 11 runs produce 50% of the mean peak dose, with these individual contributions of the highest contributing runs: 10.2%, 5.3%, 5.1%, 5.0%, 4.3%, 4.1%, 3.8%,… In the 25,000 runs calculation 17 runs produce 50% of the mean peak dose, with these individual contributions of the highest contributing runs: 8.7%, 8.2%, 6.9%, 4.5%, 2.9%, 2.3%, 2,0%,…. Figure 4.36 presents the CSM plot for the peak dose due to Ni59 (output variable) for the non-relevant parameters that produce the highest value of D_m in the 5,000 and 25,000 runs calculations (KDBESN and KDRTH, respectively).

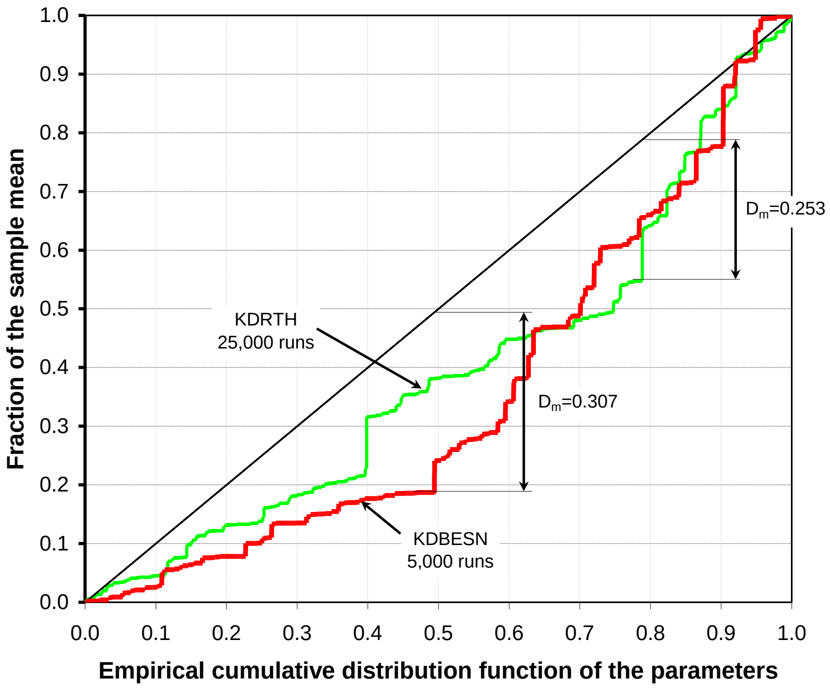


Figure 4.36 – CSM plot for “worst” non-relevant parameters and the peak dose due to Ni59 (output variable) in the 5,000 and 25,000 runs calculations.

Comparison of the tables of the test statistics with 5,000 and 25,000 runs shows that the decrease in the threshold allows identifying as relevant new parameters for which the test statistic is smaller than the threshold value in the 5,000 runs calculation and smaller than the threshold value in the 25,000 runs calculation. For the parameters already identified as relevant in the 5,000 runs calculation, the test statistic takes similar values in the 5,000 and 25,000 runs calculations.

Table 4.27 – Test statistic for CSM of the peak dose due to C14, Ni59, Sn126 and Tc99 for different parameters (25,000 runs).

2500 runs	C14	Ni59	Sn126	Tc99
Waste (3)				
R1000	--	--	0.084	--
Canisters (3)				
FAIL	0.122	--	--	--
SLOPE	0.106	--	--	--
Near field (6)				
WATER NF	0.360	0.384	0.454	0.450
SOLX		0.309	0.099	0.285
KDBEX	0.378		0.452	
TT NF (1/TT NF)	0.496	0.410	0.586	0.420
MF NF	--	0.415	0.368	0.507
Water flow (4)				
WATER TT0	0.756	0.791	0.510	0.604
KINEM POR	0.291		0.124	
Granite matrix (6)				
FWA	0.275	0.642	0.332	0.433
THICKM	0.254	0.505	0.217	0.298
MATCOR	0.280	0.463	0.261	0.283
KDRX	0.202	0.422	0.304	0.334
TT FF (1/TT FF)	0.855	0.958	0.754	0.852
TT NF+TT FF	0.818	0.918	0.698	0.760
DecayT/(TT NF+TT FF)	0.873	0.959	0.832	0.869
MF NF*DecayT	--	0.964	0.789	0.878
Empirical threshold	0.098	0.253	0.078	0.080

Table 4.28 – Test statistic for CSM of the peak dose due to Se79 and Pd107 for different parameters (25,000 runs).

25,000 runs	Se79	Pd107
Waste (3)		
Canisters (3)		
Near field (6)		
POR ANIONS	0.033	--
WATER NF	0.432	0.447
SOLX	0.459	0.571

DIFFX	0.039	--
TT NF (1/TT NF)	0.385	0.355
MF NF	0.589	0.670
Water flow(4)		
WATER TT0	0.120	0.078
KINEM POR	0.046	0.034
Granite matrix (6)		
FWA	0.110	0.110
THICKM	0.079	0.096
MATCOR	0.077	0.086
KDRX	0.079	0.148
TT FF (1/TT FF)	0.233	0.250
TT NF+TT FF	0.322	0.420
MF NF*DecayT	0.621	0.701
Empirical threshold	0.028	0.032

Table 4.29 – Test statistic for CSM of the peak dose due to the 4 actinide series for different parameters (25,000 runs).

25000 runs	4n series (U236)	4n+1 series (Np237)	4n+2 series (U238)	4n+3 series (U235)
Waste (3)				
Canisters (3)				
SLOPE	--	0.107	--	--
Near field (6)				
WATER NF	0.449	0.415	0.463	0.455
SOLNP		0.528		
SOLU	0.545	--	0.613	0.488
Water flow (4)				
WATER TT0	0.464	0.617	0.505	0.566
TT NF (1/TT NF)	0.433	0.417	0.447	0.435
MF NF	0.643	0.639	0.703	0.597
Granite matrix (6)				
FWA	0.356	0.476	0.352	0.412
THICKM	0.248	0.345	0.276	0.329
MATCOR	0.317	0.343	0.332	0.360
KDRPA	--	--		0.270
KDRRA	--	--	0.159	--
KADRTH	--	0.264	0.126	--
KDRU	0.166	--		--
TT FF (1/TT NF)	0.748	0.823	0.746	0.749
TT NF+TT FF	0.762	0.725	0.766	0.772
DecayT/TT NF+TT FF	--	0.846	--	--
MF NF*DecayT	--	0.881	--	--

Empirical threshold	0.093	0.105	0.107	0.151

4.6 Ranking of parameters

Comparison of the tables generated in the previous sections shows that for a given radionuclide the parameters identified as relevant are nearly the same for the different statistics.

The values of the different statistics for all the random input parameters and derived parameters for the peak dose due to I129 and Se79 are summarised in Table 4.30 and Table 4.32 respectively. On the basis of these tables a ranking of importance of the parameters is created for each statistic and the results are shown in Table 4.31 for I129 and Table 4.33 for Se79.

The 6 different rankings obtained for each radionuclide are quite similar. The parameters with greater influence on the peak dose are "derived parameters" both for I129 and Se79.

Having obtained similar rankings of importance of the parameters using different statistics confirms the robustness of the results obtained in the sensitivity analysis. Although each statistic provides a particular information on the parameters, the same handful of parameters controls the doses due to a given radionuclide in all the cases.

Table 4.30 – Correlation coefficients and Monte Carlo filtering and CSM statistics for random and derived parameters vs. I129 peak doses.

	Values				Ranks	
	CC SRC	Smirnov 10%-90%	t* 10%-90%	CSM	RCC SRRC	M-W* 10%-90%
Waste(3)						
R1000	0.131	0.081	0.055	0.044	0.153	0.101
GAPI	0.403	0.406	0.278	0.125	0.394	0.516
Canisters (3)						
MEAN	-0.071	0.104	0.066	0.018	-0.046	-0.126
SLOPE	0.037	0.059	0.033	--	--	0.064
Near field (6)						
POR ANIONS BUFFER					-0.023	
WATER NF	0.400	0.438	0.270	0.156	0.558	0.541
KDBEI	-0.221	0.162	0.105	0.075	-0.252	-0.201
TT NF	-0.495	0.505	0.208	0.176	-0.615	-0.593
Water flow (4)						
WATER TT0	-0.380	0.451	0.317	0.118	-0.334	-0.554
KINEM POR	-0.297	0.233	0.136	0.088	-0.340	-0.310
PECLET	0.122	0.086	0.059	0.036	0.131	0.108
Granite matrix (6)						
FWA	-0.057	0.073	0.048	0.018	-0.046	-0.095
THICKM	-0.033	--	0.026	--	-0.032	-0.050
TT FF	-0.465	0.574	0.278	0.157	-0.502	-0.689
TT NF+TT FF	-0.679	0.759	0.350	0.235	-0.862	-0.877
GAP/(TT NF+TT FF)	0.780	0.850	0.895	0.257	0.952	0.942

Table 4.31 – Rankings of importance of the main parameters affecting I129 peak dose.

	Values				Ranks	
	CC SRC	Smirnov 10%-90%	t* 10%-90%	CSM	RCC SRRC	M-W* 10%-90%
Waste(3)						
R1000	10	12	12	10	10	12
GAPI	6	7	5	6	6	7
Canisters (3)						
MEAN	12	10	10	12	12	10
SLOPE						
Near field (6)						
POR ANIONS BUFFER						
WATER NF	5	6	6	5	4	6
KDBEI	9	9	9	9	9	9
TT NF	3	4	7	3	3	4
Water flow (4)						
WATER TT0	7	5	3	7	8	5
KINEM POR	8	8	8	8	7	8
PECLET	10	11	11	11	11	11
Granite matrix (6)						
FWA						
THICKM						
TT FF	4	3	4	4	5	3
TT NF+TT FF	2	2	2	2	2	2
GAP/(TT NF+TT FF)	1	1	1	1	1	1

Table 4.32 – Correlation coefficients and Monte Carlo filtering and CSM statistics for random and derived parameters vs. Se79 peak dose.

	Values				Ranks	
	CC SRC	Smirnov 10%-90%	t* 10%-90%	CSM	RCC SRRC	M-W* 10%-90%
Waste (3)						
Canisters (3)						
Near field (6)						
POR ANIONS BUFFER	0.028			0.033		
WATER NF	0.437	0.586	0.420	0.432	0.582	0.707
SOLSE	0.492	0.625	0.488	0.459	0.627	0.753
DIFFSE	0.033			0.039	0.028	
MF NF	0.837	0.910	0.967	0.589	0.863	0.961
Water flow (4)						
WATER TT0	-0.112	0.139	0.092	0.120	-0.178	-0.169
KINEM POR	-0.041	0.046	0.028	0.046	-0.039	-0.047
Granite matrix (6)						
FWA	-0.099	0.130	0.087	0.110	-0.218	-0.117
THICKM	-0.075	0.093	0.064	0.079	-0.160	-0.118
MATCOR	-0.073	0.087	0.058	0.077	-0.154	-0.111

KDRSE	-0.073	0.098	0.065	0.079	-0.157	-0.122
DecayT	0.220	0.272	0.190	0.233	0.395	0.355
MF NF*DecayT	0.974	0.968	1.212	0.621	0.995	0.996

Table 4.33 – Rankings of importance of the main parameters affecting Se79 peak dose.

	Values				Ranks	
	CC SRC	Smirnov 10%-90%	t* 10%-90%	CSM	RCC SRRC	M-W* 10%-90%
Waste (3)						
Canisters (3)						
Near field (6)						
POR ANIONS BUFFER	12			12		
WATER NF	4	4	4	4	4	4
SOLSE	3	3	3	3	3	3
DIFFSE						
MF NF	2	2	2	2	2	2
Water flow (4)						
WATER TT0	6	6	6	6	7	6
KINEM POR	11	11	11	11	11	11
Granite matrix (6)						
FWA	7	7	7	7	6	9
THICKM	8	9	9	9	8	8
MATCOR	10	10	10	10	10	10
KDRSE	9	8	8	8	9	7
DecayT	5	5	5	5	5	5
MF NF*DecayT	1	1	1	1	1	1

4.7 Effect on the sensitivity analysis of including a stochastic Biosphere

In many Safety Assessments exercises (including those performed by ENRESA) the Biosphere is included in the calculations through 2 parameters:

- a flow of water where the radionuclides released from the host formation are dissolved (BIOFLOW), and
- a set of radionuclide specific Dose Conversion Factors (DCF) that transform the concentration in the water used by the critical group into dose.

The dose due to a radionuclide is calculated in this way:

$$\text{Dose}(\text{Sv / year}) = \text{Activity Flux from Geosphere} \cdot \frac{\text{DCF}}{\text{BIOFLOW}}$$

In the calculations performed in sections 4.1 to 4.6 DCF and BIOFLOW were constants, not probability distributions.

In order to explore the effects of an stochastic treatment of the Biosphere on the sensitivity analysis of a repository, a new 25,000 runs calculation was done considering a factor 4 of uncertainty in BIOFLOW and a factor 10 of uncertainty in all the DCF. BIOFLOW was represented as the product of a log-uniform distribution between 0.5 and 2 ($BIOFLOW_R$) times the value used in the previous calculations (10^6 m^3/yr) and each DCF was represented by a log-uniform distribution between 0.3 and 3 (DCF_R) times the constant DCFs used in the base calculations.

In reality it was not necessary to rerun the safety assessment model. Instead, DCF_R and $BIOFLOW_R$ were sampled and the quotient $DCF_R/BIOFLOW_R$ was calculated 25,000 times. The doses obtained in run **j** of the calculation with constant biosphere were multiplied by the value of $DCF_R/BIOFLOW_R$ obtained in run **j** of the stochastic calculation with constant Biosphere factors to obtain the doses with stochastic biosphere.

4.7.1 Effect of the stochastic biosphere on the variance of the peak doses

The effect of the stochastic treatment of the biosphere proposed is to widen the range of variation of the peak doses compared with the case of constant biosphere. The width is quantified by the variance of the distribution of peak doses and since the peak dose of a given radionuclide typically spans over several orders of magnitude it is appropriate to take logarithms to discuss the variance in terms of orders of magnitude.

The relationship between the peak dose with the stochastic biospheres (SB) and the constant biosphere (CB) in run j is:

$$\text{Peak Dose}_{SBj} = \text{Peak Dose}_{CBj} \cdot \left(\frac{DCF_R}{BIOFLOW_R} \right)_j$$

Taking logarithms transforms the product into a sum. Using that the variance of the sum of two independent stochastic parameters is the sum of the variances of the two parameters, it is obtained that:

$$\mathrm{Var}[\log(\text{Peak Dose}_{SB})] = \mathrm{Var}[\log(\text{Peak Dose}_{CB})] + \mathrm{Var}\left[\log\left(\frac{DCF_R}{BIOFLOW_R}\right)\right]$$

With the pdf´s adopted for $DCF_R/BIOFLOW_R$ the variance added by the stochastic biosphere is 0.115.

$$\mathrm{Var}\left[\log\left(\frac{DCF_R}{BIOFLOW_R}\right)\right] = 0.115$$

The variance of the distributions of the (decimal) logarithm of the peak doses for all the non-zero runs in the 25,000 runs have been calculated with the stochastic and the constant biospheres (see Table 4.34). Column "Δ Variance" of the table is calculated as the variance with stochastic biosphere minus the variance with constant biosphere, and is very similar to the predicted value of 0.115. The coincidence is not complete due to statistical variations.

Table 4.34 – Variance of the distributions of the logarithm of the peak doses with constant and stochastic biospheres.

	$Var[\log(Peak\ \ Dose)]$		Δ Variance
	Constant biosphere	Stochastic biosphere	
Total	0.090	0.205	0.115
C14	23.530	23.625	0.095
Cl36	0.112	0.227	0.115
Cs135	33.832	33.912	0.080
I129	0.088	0.203	0.115
Ni59	33.559	33.642	0.083
Pd107	4.676	4.779	0.103
Se79	1.056	1.165	0.109
Sn126	24.349	24.439	0.090
Tc99	30.367	30.469	0.102
4n	40.581	40.671	0.090
4n+1	42.708	42.867	0.159
4n+2	48.851	48.926	0.075
4n+3	44.077	44.150	0.073

Table 4.34 shows that using the stochastic biosphere defined previously produces a significant increase of the variance of the peak dose (higher than a factor 2) only for the total dose, Cl36 and I129. For Se79 the increase is roughly a 10% and for the rest of radionuclides the increase in the variance of the peak dose is negligible.

It is reasonable to expect that the effects of the stochastic biosphere on the sensitivity analysis will be important for the total, Cl36 and I129 peak doses, small for Se79 peak doses and very small for the peak doses due to Pd107 and the rest of radionuclides.

4.7.2 Correlation coefficients

Table 4.35 presents the results of the correlation analyses for I129 peak dose using stochastic and constant biospheres. An important decrease in the correlation coefficients of all the parameters is observed. For the parameters with high correlation coefficients the CCs with stochastic biosphere are around 0.57 times the values with constant biosphere, while for RCCs this factor is 0.64. For less correlated parameters the reduction factor presents more dispersion.

Table 4.36 presents the results of the correlation analyses for Se79 peak dose using stochastic and constant biospheres. A significant decrease in the correlation coefficients of the parameters is observed. For the parameters with high correlation coefficients the CCs with stochastic biosphere are around 0.75 times the values with constant biosphere, while for RCCs this factor is 0.94. For less correlated parameters the reduction factor presents more dispersion.

The stochastic biosphere produces a higher decrease in the CCs than in the RCCs. The decreases in the RCCs due to the stochastic biosphere are in good agreement with the predictions done in 4.7.2.1.

Finally, Table 4.37 presents the results of the correlation analyses for Cs135 peak dose using stochastic and constant biospheres. As shown in Table 4.34, the stochastic biosphere produces a very small relative increase in the variance of the logarithm of the peak dose due to Cs135 and the RCCs with the stochastic and constant biosphere are nearly identical. By the contrary, for the peak dose due to Cs135 the RCCs with the stochastic biosphere are about 0.77 times the CCs with the constant biosphere. Similar results are obtained for other radionuclide with high values for the variance of the peak dose: the RCCs with stochastic and constant biospheres are nearly identical while the CCs with the stochastic biosphere are 0.75-0.9 times the CCs with the constant biosphere.

The CCs and RCCs of the biosphere parameters BIOFLOW and DCF and the derived parameter DCF/BIOFLOW take high values for the 3 radionuclides. For I129, DCF/BIOFLOW is the parameter with the greatest correlation coefficients and explains 46% of the variance of the peak dose due in values and 56% of the variance in ranks.

Table 4.35 – Correlation and regression analyses for I129 peak dose with constant and stochastic Biospheres.

	Constant Biosphere		Stochastic Biosphere		Ratio Stochastic/Constant	
	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC
Waste(3)						
R1000	0.131	0.153	0.095	0.124	0.725	0.810
GAPX	0.403	0.394	0.236	0.266	0.586	0.675
Canisters (3)						
MEAN	-0.071	-0.046	-0.034	-0.025	0.479	0.543
SLOPE	0.037		0.022		0.595	
Near field (6)						
POR ANIONS BUFFER		-0.023		-0.013		
WATER NF	0.400	0.558	0.228	0.360	0.570	0.645
KDBEX	-0.221	-0.252	-0.123	-0.164		
TT NF	-0.495	-0.615	-0.286	-0.395	0.578	0.642
Water flow (4)						
WATER TT0	-0.380	-0.334	-0.210	-0.210	0.553	0.629
KINEM POR	-0.297	-0.340	-0.179	-0.223	0.603	0.656
PECLET	0.122	0.131	0.080	0.085	0.656	0.649
Granite matrix (6)						
FWA	-0.057	-0.046	-0.037	-0.028	0.649	0.609
THICKM	-0.033	-0.032	-0.018	-0.014	0.545	0.438
TT FF	-0.465	-0.502	-0.265	-0.320	0.570	0.637
Biosphere (2)						
BIOFLOW			-0.314	-0.375		
DCF			0.546	0.649		
DCF/BIOFLOW			0.680	0.750		
TT NF+TT FF	-0.679	-0.862	-0.389	-0.552	0.573	0.640
GAP/(TT NF+TT FF)	0.780	0.952	0.436	0.610	0.559	0.641

Table 4.36 – Correlation and regression analyses for Se79 peak dose with constant and stochastic Biospheres.

	Constant Biosphere		Stochastic Biosphere		Ratio Stochastic/Constant	
	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC
Near field (6)						
POR ANIONS BUFFER	0.028		0.029		1.036	
WATER NF	0.437	0.582	0.337	0.545	0.771	0.936
SOLX	0.492	0.627	0.381	0.584	0.774	0.931
DIFFSE	0.033	0.028	0.020	0.030	0.606	1.071
MF NF	0.837	0.863	0.655	0.803	0.783	0.930
Water flow (4)						
WATER TT0	-0.112	-0.178	-0.078	-0.168	0.696	0.944
KINEM POR	-0.041	-0.039	-0.035	-0.040	0.854	1.026
PECLET						
Granite matrix (6)						
FWA	-0.099	-0.218	-0.074	-0.210	0.747	0.963
THICKM	-0.075	-0.160	-0.055	-0.152	0.733	0.950
MATCOR	-0.073	-0.154	-0.052	-0.143	0.712	0.929
KDRSE	-0.073	-0.157	-0.053	-0.145	0.726	0.924
DecayT	0.220	0.395	0.164	0.373	0.745	0.944
Biosphere (2)						
BIOFLOW			-0.110	-0.164		
DCF			0.204	0.284		
DCF/BIOFLOW			0.245	0.325		
MF NF*DecayT	0.974	0.995	0.755	0.929	0.775	0.934

Table 4.37 – Correlation and regression analyses for Cs135 peak dose with constant and stochastic Biospheres.

	Constant Biosphere		Stochastic Biosphere		Ratio Stochastic/Constant	
	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC
Waste(3)						
R1000	0.092	0.049	0.075	0.049	0.815	1.000
GAPX	0.030		0.019		0.633	
Canisters (3)						
Near field (6)						
WATER NF	0.189	0.102	0.138	0.097	0.730	0.951
KDBEX	-0.070	-0.032	-0.057	-0.031	0.814	0.969
TT NF	-0.151	-0.107	-0.116	-0.102	0.768	0.953
1/TT NF	0.185	0.107	0.140	0.102	0.757	0.953
Water flow (4)						
WATER TT0	-0.295	-0.438	-0.224	-0.431	0.759	0.984
PECLET		-0.189		-0.190		1.005
Granite matrix (6)						

FWA	-0.178	-0.526	-0.139	-0.523	0.781	0.994
THICKM	-0.161	-0.365	-0.125	-0.363	0.776	0.995
MATCOR	-0.170	-0.378	-0.127	-0.373	0.747	0.987
KDRX	-0.119	-0.270	-0.094	-0.268	0.790	0.993
TT FF	-0.125	-0.958	-0.096	-0.950	0.768	0.992
1/TT FF	0.523	0.958	0.401	0.950	0.767	0.992
DecayT	0.685	0.958	0.527	0.950	0.769	0.992
Biosphere (2)						
BIOFLOW			-0.079	-0.044		
DCF			0.152	0.081		
DCF/BIOFLOW			0.182	0.091		
TT NF+TT FF	-0.132	-0.955	-0.101	-0.947	0.765	0.992
1/(TT NF+TT FF)	0.928	0.955	0.709	0.947	0.764	0.992
DecayT/(TT NF+TT FF)	0.906	0.965	0.693	0.957	0.765	0.992

4.7.2.1 Prediction of the RCCs with the stochastic biosphere

The reduction in the RCCs produced by the stochastic biosphere can be explained on the basis of the increase in the variance of the peak dose. As shown in section 4.1.3, the correlation coefficients and the coefficient of determination (R^2) obtained with the transformations into logarithms and ranks are very similar. After making the transformation into logarithms, Table 4.34 shows that for I129 57% (0.115/0.203) of the total variance corresponds to the biosphere parameters and 43% (0.088/0.203) of the variance corresponds to the other parameters. For Se79 10% of the variance corresponds to biosphere parameters and 90% to the other parameters.

For I129, the coefficient of determination due to the non-biosphere parameters with the stochastic biosphere (R^2_{SB}) will be 0.43 times the coefficient of determination with the constant biosphere (R^2_{CB}). Assuming that the RCCs with the stochastic biosphere (RCC_{SB}) are proportional to the correlation coefficients with constant biosphere (RCC_{CB}), it is obtained that

$$RCC_{SB}(I129) = \sqrt{0.43} \cdot RCC_{CB}(I129) = 0.66 \cdot RCC_{CB}(I129)$$

Similarly, for Se79 it is obtained that

$$RCC_{SB}(Se79) = \sqrt{0.10} \cdot RCC_{CB}(Se79) = 0.95 \cdot RCC_{CB}(I129)$$

Values in the last column of Table 4.38 and Table 4.39 are very close to the predicted values (0.66 for I129 and 0.95 for Se79), confirming the validity of the discussion of the effects of the stochastic biosphere in terms of the variance of the distribution of the logarithm of the peak doses.

4.7.3 Monte Carlo filtering (Smirnov. Mann-Whitney and t tests)

Table 4.38 to Table 4.40 present the results of the Monte Carlo filtering analyses with constant and stochastic biospheres for I129, Se79 and Cs135. The values of the statistics with the stochastic biosphere are smaller than in the case of the constant biosphere for the 3 radionuclides.

The Monte Carlo filtering statistics of the biosphere parameters BIOFLOW and DCF and the derived parameter DCF/BIOFLOW take high values for the 3 radionuclides. For I129 DCF/BIOFLOW is the parameter with the greatest values for these statistics.

Table 4.38 – Monte Carlo filtering for I129 peak dose (10% of highest doses vs. remaining 90%) with constant and stochastic Biospheres.

	Constant Biosphere			Stochastic Biosphere			Ratio Stochastic/Constant		
	Smir	M-W*	T*	Smir	M-W*	t*	Smir	M-W*	t*
Waste(3)									
R1000	0.084	0.104	0.057	0.100	0.117	0.059	1.190	1.125	1.035
GAPI	0.429	0.548	0.297	0.258	0.331	0.175	0.601	0.604	0.589
Canisters (3)									
MEAN	0.104	-0.126	0.066	0.053	-0.056	0.029	0.510	0.444	0.439
SLOPE	0.062	0.068	0.035	0.036	0.037	0.019	0.581	0.544	0.543
Near field (6)									
WATER NF	0.428	0.530	0.261	0.314	0.392	0.170	0.734	0.740	0.651
KDBEI	0.177	-0.225	0.118	0.126	-0.154	0.080	0.712	0.684	0.678
TT NF	0.504	-0.594	0.208	0.357	-0.430	0.182	0.708	0.724	0.875
Water flow (4)									
WATER TT0	0.470	-0.542	0.308	0.214	-0.277	0.152	0.455	0.511	0.494
KINEM POR	0.251	-0.335	0.149	0.207	-0.271	0.129	0.825	0.809	0.866
PECLET	0.098	0.124	0.067	0.096	0.122	0.064	0.980	0.984	0.955
Granite matrix (6)									
FWA	0.056	-0.063	0.038	0.049	-0.049	0.030	0.875	0.778	0.789
THICKM									
TT FF	0.577	-0.690	0.280	0.323	-0.417	0.190	0.560	0.604	0.679
Biosphere (2)									
BIOFLOW				0.345	-0.446	0.225			
DCF				0.555	0.694	0.426			
DCF/BIOFLOW				0.659	0.806	0.590			
TT NF+TT FF	0.763	-0.881	0.352	0.486	-0.620	0.267	0.637	0.704	0.759
GAP/(TT NF+TT FF)	0.865	0.955	0.904	0.553	0.688	0.342	0.639	0.720	0.378

Table 4.39 – Monte Carlo filtering for Se79 peak doses (10% of highest doses vs. remaining 90%) with constant and stochastic Biospheres.

	Constant Biosphere			Stochastic Biosphere			Ratio Stochastic/Constant		
	Smir	M-W	T	Smir	M-W	T	Smir	M-W	T
Near field (6)									
WATER NF	0.586	0.707	0.420	0.524	0.651	0.349	0.894	0.921	0.831
SOLSE	0.625	0.753	0.488	0.561	0.698	0.434	0.898	0.927	0.889
MF NF	0.910	0.961	0.967	0.784	0.905	0.762	0.862	0.942	0.788
Water flow (4)									
WATER TT0	0.139	-0.169	0.092	0.112	-0.139	0.076	0.806	0.822	0.826

KINEM POR	0.046	-0.047	0.028	0.044	-0.047	0.025	0.957	1.000	0.893
Granite matrix (6)									
FWA	0.130	-0.177	0.087	0.130	-0.172	0.080	1.000	0.972	0.920
THICKM	0.093	-0.118	0.064	0.071	-0.099	0.052	0.763	0.839	0.813
MATCOR	0.087	-0.111	0.058	0.083	-0.098	0.052	0.954	0.883	0.897
KDRSE	0.098	-0.122	0.065	0.083	-0.095	0.050	0.847	0.779	0.769
TT FF	0.272	-0.355	0.101	0.239	-0.312	0.097	0.879	0.879	0.960
Biosphere (2)									
BIOFLOW				0.162	-0.214	0.108			
DCF				0.279	0.366	0.192			
DCF/BIOFLOW				0.311	0.420	0.228			
TT NF+TT FF	0.392	-0.529	0.116	0.356	-0.477	0.111	0.908	0.902	0.957
MF NF*DecayT	0.968	0.996	1.212	0.837	0.943	0.905	0.865	0.947	0.747

Table 4.40 – Monte Carlo filtering for Cs135 peak doses (10% of highest doses vs. remaining 90%) with constant and stochastic Biospheres.

	Constant Biosphere			Stochastic Biosphere			Ratio Stochastic/Constant		
	Smir	M-W*	t*	Smir	M-W*	t*	Smir	M-W*	t*
Waste(3)									
R1000	0.132	0.170	0.089	0.132	0.168	0.087	1.000	0.988	0.978
GAPX		0.038	0.020	0.038	0.033	0.017		0.868	0.850
Canisters (3)									
Near field (6)									
WATER NF	0.323	0.393	0.164	0.276	0.341	0.144	0.854	0.868	0.878
KDBEX	0.082	-0.111	0.059	0.080	-0.106	0.055	0.976	0.955	0.932
TT NF	0.332	-0.412	0.180	0.281	-0.358	0.160	0.846	0.869	0.889
1/TT NF	0.332	0.412	0.146	0.281	0.358	0.131	0.846	0.869	0.897
Water flow (4)									
WATER TT0	0.487	-0.582	0.335	0.463	-0.550	0.316	0.951	0.945	0.943
Granite matrix (6)									
FWA	0.381	-0.507	0.194	0.374	-0.500	0.192	0.982	0.986	0.990
THICKM	0.284	-0.361	0.174	0.271	-0.349	0.169	0.954	0.967	0.971
MATCOR	0.301	-0.374	0.181	0.281	0.362	0.174	0.934	-0.968	0.961
KDRX	0.196	-0.268	0.126	0.186	-0.259	0.124	0.949	0.966	0.984
TT FF	0.874	-0.938	0.139	0.818	-0.914	0.139	0.936	0.974	1.000
1/TT FF	0.874	0.938	0.488	0.818	0.914	0.451	0.936	0.974	0.924
DecayT	0.874	0.938	1.000	0.818	0.914	0.877	0.936	0.974	0.877
Biosphere (2)									
BIOFLOW				0.131	-0.166	0.083			
DCF				0.229	0.293	0.151			
DCF/BIOFLOW				0.247	0.334	0.173			
TT NF+TT FF	0.916	-0.986	0.147	0.844	-0.914	0.146	0.921	0.927	0.993
1/(TT NF+TT FF)	0.916	0.986	0.986	0.844	0.914	0.857	0.921	0.927	0.869
DecayT/(TT NF+TT FF)	0.913	0.984	0.742	0.846	0.914	0.663	0.927	0.929	0.894

4.7.4 Contribution to the sample mean (CSM)

Table 4.41 and Table 4.42 present the values of the test statistic for CSM of the peak doses due to I129 and Se79 for constant and stochastic biospheres. In this case the use of a stochastic biosphere has no significant effects: for parameters with high values of the statistic the value remains nearly unchanged and only for parameters with low values of the statistic small differences are observed, but without any clear trend of increase or decrease.

Table 4.41 – Test statistic for CSM of the peak dose due to I129 for constant and stochastic Biospheres (25,000 runs).

	Constant Biosphere	Stochastic Biosphere	Ratio Stochastic/Constant
Waste (3)			
R1000	0.039	0.050	1.282
GAPX	0.118	0.125	1.059
Canisters (3)			
MEAN	0.022	0.021	0.955
Near field (6)			
WATER NF	0.154	0.156	1.013
KDBE	0.067	0.066	0.985
TT NF (1/TT NF)	0.172	0.173	1.006
Water flow (4)			
WATER TT0	0.109	0.104	0.954
KINEM POR	0.089	0.097	1.090
PECLET	0.036	0.041	1.139
Granite matrix (6)			
FWA	0.016	0.019	1.188
TT FF (1/TT FF)	0.151	0.149	0.987
Biosphere (2)			
BIOFLOW		0.168	
DCF		0.270	
DCF/BIOFLOW		0.301	
TT NF+TT FF	0.230	0.231	1.004
GAP/(TT NF+TT FF)	0.253	0.254	1.004

Table 4.42 – Test statistic for CSM of the peak dose due to Se79 for constant and stochastic Biospheres (25,000 runs).

	Constant Biosphere	Stochastic Biosphere	Ratio Stochastic/Constant
Waste (3)			
Canisters (3)			
Near field (6)			
POR ANIONS	0.033	0.041	1.242
WATER NF	0.432	0.433	1.002
SOLSE	0.459	0.460	1.002
DIFFSE	0.039	0.032	0.821
TT NF (1/TT NF)	0.385	0.387	1.005
MF NF	0.589	0.591	1.003

Water flow(4)			
WATER TT0	0.120	0.109	0.908
KINEM POR	0.046	0.057	1.239
Granite matrix (6)			
FWA	0.110	0.109	0.991
THICKM	0.079	0.079	1.000
MATCOR	0.077	0.072	0.935
KDRSE	0.079	0.073	0.924
TT FF (1/TT FF)	0.233	0.225	0.966
Biosphere (2)			
BIOFLOW		0.159	
DCF		0.272	
DCF/BIOFLOW		0.297	
TT NF+TT FF	0.322	0.315	0.978
MF NF*DecayT	0.621	0.621	1.000

4.7.5 Conclusions on the stochastic biosphere

The effects of an stochastic biosphere in the sensitivity analysis of Safety Assessment calculations can be broadly predicted on the basis of the additional variance introduced by the stochastic biosphere (section 4.7.1).

The values of the statistics for the random input parameters of the biosphere are positively correlated with the ratio of the variance introduced by the stochastic biosphere divided by the original variance of the peak dose with a constant biosphere:

$$\text{Ratio (Var)} = \frac{\text{Var}\left[\log\left(\frac{\text{DCF}_R}{\text{BIOFLOW}_R}\right)\right]}{\text{Var}\left[\log(\text{Peak Dose}_{CB})\right]}$$

For I129 peak doses (Ratio(Var)=1.31) the biosphere parameters have the highest values for the different statistics, and DCF/BIOFLOW would lead the ranking of parameters. For other radionuclides with much smaller values of Ratio(Var), such as Cs135, the values of the statistics for the biosphere parameters remain significant but are not among the most important parameters.

For the stochastic parameters not related with the biosphere the use of a stochastic biosphere produces a decrease in the CCs, RCCs and Monte Carlo filtering statistics. The higher the value of Ratio(Var) for a given radionuclide, the greater will be the reduction in the statistics. Although the numerical values change the relative importance of the different parameters for a given statistic remains nearly unchanged compared with the case with constant biosphere. By the contrary the values of the CSM statistic are not changed by the use of an stochastic biosphere.

5 Conclusions and recommendations

The different statistics used in the sensitivity analysis identify the same set of relevant parameters with a similar ranking of importance. This agreement confirms the robustness of the results obtained in the sensitivity analysis. The different sensitivity analysis methods are complementary, and each one provides specific information on the parameters,.

The capability to identify small effects in the stochastic model increases with the number of runs, as expected. The threshold value above which an statistic is considered statistically significant decreases with the number of runs n following the “rule of thumb”:

$$\text{threshold} \, \alpha \, \frac{1}{\sqrt{n}}$$

The sampling method used (SRS or LHS) makes no difference in the correlation and regression analysis. No formal comparison of both sampling schemes has been done for the rest of statistics. SRS is preferred from a practical perspective because it allows performing several parallel stochastic calculations using different computers (or cores of a multi-core PC) and then append the runs to produce a calculation with many runs (25,000 in our case). With LHS this approach is not possible and the 25,000 runs must be calculated in the same PC.

In the time dependent regression analysis for the total dose (Figure 4.10 and especially Figure 4.11) a significant decrease in the value of the coefficient of determination (R^2) between 100,000 and 400,000 thousand years was observed. The reason for such drop was traced back to the change of sign of the regression coefficients of several important parameters with the dose due to the I129 released from the gap (see section 4.2.2). Since total doses between 100,000 and 400,000 thousand years are controlled by the I129 in the gap, the effect was transmitted to the total dose.

When a change of sign of important parameters and the resulting (and unavoidable) drop in the value of R^2 is observed in a time dependent regression analysis, the reasons of the drop should be identified and explained as done in section 4.2.2.

Use of “derived parameters” in the sensitivity analysis.

Using the experience gained in the Safety Assessment it is possible to identify some parameters that have an important effect on repository behaviour and are a combination of the random input parameters. These parameters have been called “derived parameters” in this document. Examples of “derived parameters” are the “travel time in the far field for a radionuclide” for a repository in granite and the “apparent diffusion coefficient” for a repository in clay.

“Derived parameters” should be identified at the beginning of the sensitivity analysis and treated in the same way than the random input variables through the analysis. It must be kept in mind that these “derived parameters” are correlated with some input parameters.

In the sensitivity analysis performed, “derived parameters” always got the highest values of the different statistics, showing that they have the greatest influence on the peak doses. The rankings of importance of the difference parameters presented in section 4.6 on the basis of the different statistics calculated are led by “derived parameters” in all the cases.

Use of non-relevant parameters to derive threshold values for the statistics.

The random input parameters that are known for sure that can not have any effect on the output variable can be used to derive empirical values for the minimum value of a statistic that is statistically significant (empirical thresholds). These “empirical thresholds” have been calculated for correlation coefficients, Smirnov, Mann-Whitney and t-tests and CSM plots, and are in good agreement with the “predicted thresholds” based on statistics theory.

In one case (CSM plots with 25,000 runs) “predicted thresholds” could not be calculated because MATLAB crashed, while the “empirical thresholds” were calculated with little computational effort.

These non-relevant parameters have been found very useful to establish “empirical thresholds” for different statistics. In the model used any output variable can be affected by 22 random input parameters at most and there are about 110 non-relevant random input parameters, because all the input parameters were independent.

In other models input parameters can be correlated and non-relevant parameters can be scarce. In this case “dummy” random parameters could be included in the model with pdf´s similar to the real input parameters. These dummy parameters would be sampled in the stochastic calculation but would play no role in the calculations, and could serve to derive “empirical thresholds” for statistics.

Monte Carlo filtering

It is more useful to present the results of the Monte Carlo filtering in terms of the statistics instead of the p-values. The statistics tend to the “real” value when the number of runs increases while the p-values tend to zero. The statistics are more robust parameters that allow making a ranking of the relative importance of the input parameters.

Mann-Whitney and t-test statistics were modified to make them independent of the number of runs, and in the case of Mann-Whitney statistic to have a range of variation between -1 and 1. The resulting M-W* and t* statistic were found more useful than the original M-W and t statistics to present the results of the Monte Carlo filtering analysis.

The t* statistic in general produces results consistent with the other statistics (the same parameters are identified as significant with a similar ranking), but some anomalies with the t* statistic have been observed (section 4.3.2.3). Since the t-test is based on the sampling distribution of the mean of normal variables, and none of the input random or derived parameters in the model have normal distributions, the strange results could be a consequence of the not applicability of the t-test to the model and parameters under study. This topic will be further investigated. Meanwhile, it is recommended to use only Smirnov and

Mann-Whitney tests for Monte Carlo filtering in values and ranks respectively, and avoid the use of the t statistic.

Graphic representations (cobweb plots)

Conditional cobweb plots allow identifying the parameters strongly correlated with a region of the output variable (usually the 10% of runs with highest peak doses. Graphically show the ranges of values of the different input parameters that correspond to the selected region of the output variable: high values of parameter A, low values of parameter B,…

Due to the great number of lines represented (one per run) cobweb plots can only present a few parameters in X axis, become blurred and only strong correlations can be identified visually. In addition cobweb plots can produce visual artefacts as described in section 4.4.2.

An alternative graphic representation (Figures 4.26 and 4.27) has been developed where only the mean ranks of the parameters in several subsets of the sample are represented. The figure presents the mean ranks for all the input parameters relevant for the dose due to a radionuclide in the 10% of runs with highest peak doses and in the 10% of runs with lowest peak doses. Input parameters are grouped on the basis of the part of the system that affect, and “derived parameters” are included too.

This type of plot provides much information in a condensed way: what input parameters do not affect at all the output variable, whether high values of the output variable are related to low or high values of each input (or derived) parameter and whether low values of the output variable are related to high or low values of the same input (or derived) parameters.

Contribution to the sample mean (CSM) plots

CSM plots are visual tools useful to identify if the mean value of the output variable is controlled by a particular region of values of an input parameter. In addition an statistic is calculated. Low values of the test statistic mean that the whole range of the input parameter has a similar contribution to the mean of the peak dose. By the contrary, high values of the statistic mean that the contribution of the parameter to the mean is controlled by a small range of values.

The CSM statistic allows selecting the parameters which contribution to the mean is controlled by a small region of values, and the CSM plots are used to identify exactly such region of values.

Effects of a stochastic treatment of the Biosphere

Most of the results obtained in the sensitivity analysis of a calculation done using a constant biosphere remain valid for a case with stochastic biosphere: important non-biosphere parameters and their ranking of importance remain the same. The numerical values of the statistics will decrease with the introduction of the stochastic biosphere.

The effects of an stochastic biosphere in the sensitivity analysis of Safety Assessment calculations can be broadly predicted on the basis of the additional variance introduced by the stochastic biosphere (section 4.7.1).

6 References

[1] JRC. Review of Sensitivity Analysis Methods and Experience. Task 2.1.D Techniques for Sensitivity and Uncertainty Analysis. PAMINA Project. December 2008.

[2] Enresa 2000. Evaluación del Comportamiento y de la Seguridad de un Almacenamiento de Combustible Gastado en una Formación Granítica. 49-1PP-M-15-01 Rev.0. December 2001. (In Spanish, internal document).

[3] SKB. Radiolitically induced oxidative dissolution of spent nuclear fuel. SKB Technical Report 90-08, Stockholm. 1990.

[4] Bolado, R., Castaigns, W. and Tarantola, S. (2008). Global Sensitivity Analysis: an Approach Based on the Contribution to the Sample Mean Plot. European Commission, Scientific Report EUR 23433 EN.