Sensitivity analysis techniques for the performance assessment of a radioactive waste repository

Ricardo Bolado Lavín[1], Klaus-Jürgen Röhlig[2], Dirk-Alexander Becker[3]

[1]Institute for Energy, European Commission DG-JRC, Petten, The Netherlands
[2]Technical University of Clausthal, Germany
[3]GRS-Braunschweig, Germany

Contents

1. Different interpretations of sensitivity
2. The concept of interaction
3. Screening methods
4. Global methods
   - Graphical methods
   - Monte Carlo based methods
     - Regression based methods
     - Monte Carlo filtering
     - Distribution sensitivity methods
   - Variance based methods
5. Sensitivity analysis activities under PAMINA (RTDC's 2 & 4)
6. Conclusions

Interpretations of sensitivity

Sensitivity is the study of the influence of the inputs on the outputs.

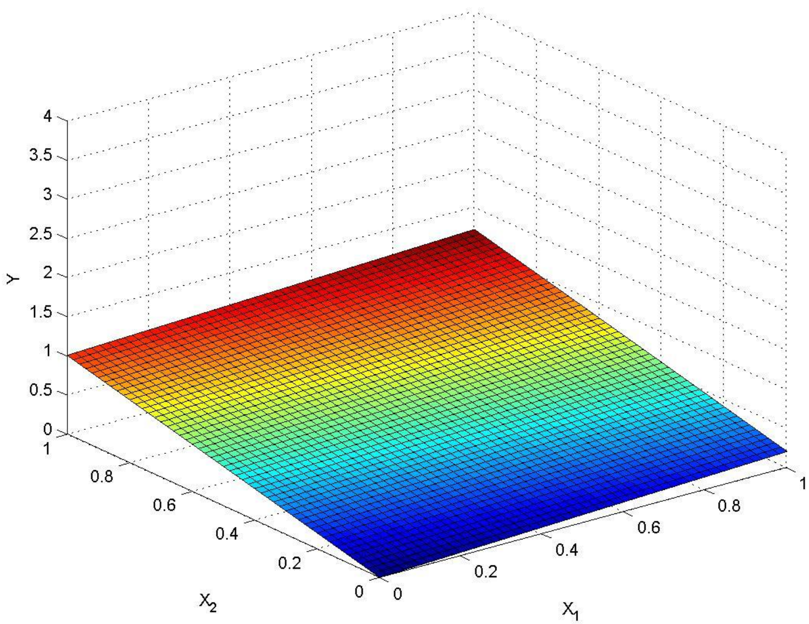
- Main problem: Under such a generic definition there may be many different interpretations.
  - Output variable response to an increment in some of the inputs (non-probabilistic interpretation).
  - Correlation between inputs and outputs
  - Monotonic relation between inputs and outputs
  - A more complex polynomial (or non-polynomial) relation
  - Specific relations between different regions of input parameters and output variables
  - Output distribution changes as a result of input distribution changes (distribution sensitivity techniques)
  - Fractional contribution to the output variance (variance based techniques)

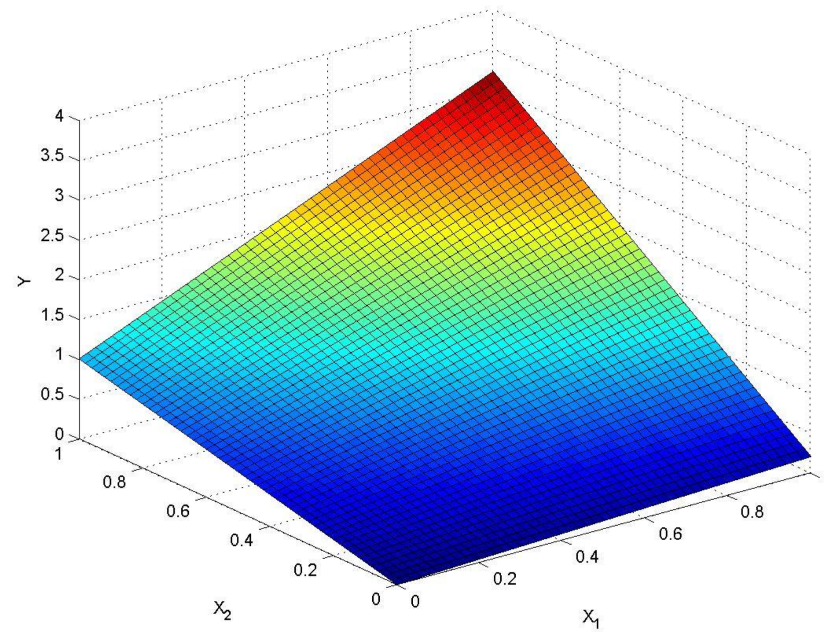
The concept of interaction

Any multivariate integrable function with support in the n-dimensional unit hypercube admits a unique decomposition in orthogonal terms (Sobol's HDMR):

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{i=n} f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) + \sum_{i<j<k} f_{ijk}(x_i, x_j, x_k) + \ldots + f_{12\ldots n}(\mathbf{x})$$

If our model has only two input parameters: $f(\mathbf{x}) = f_1(x_1) + f(x_2) + f(x_1, x_2)$



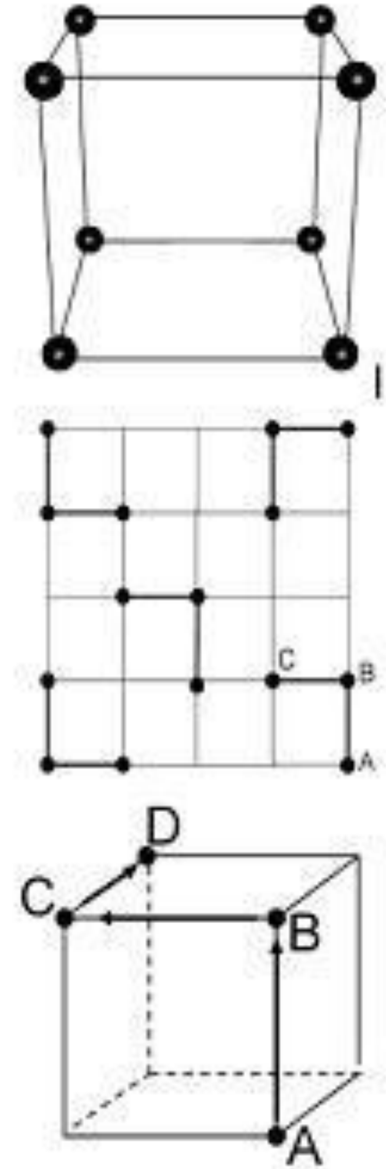


Screening methods

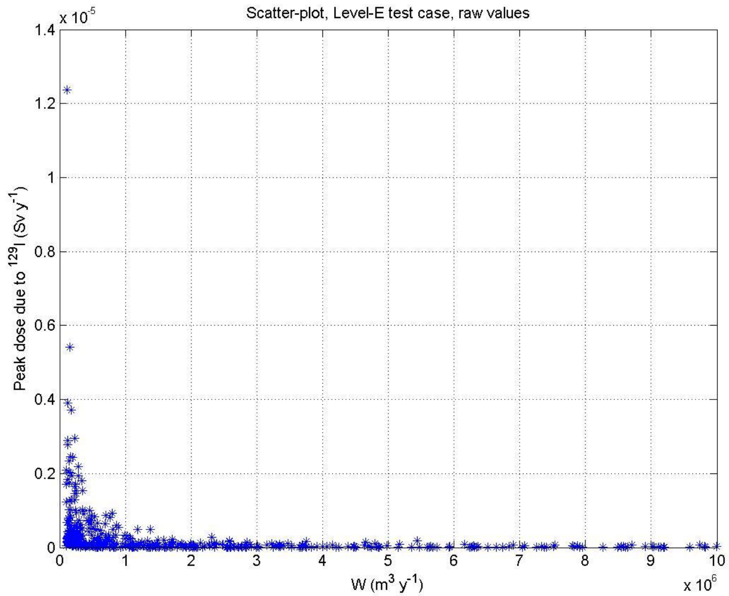
Target: Identify non-relevant input parameters from the point of view of the functional relation between inputs and outputs (only ranges of input parameters are taken into account, pdf's are ignored) ➔ screen them out

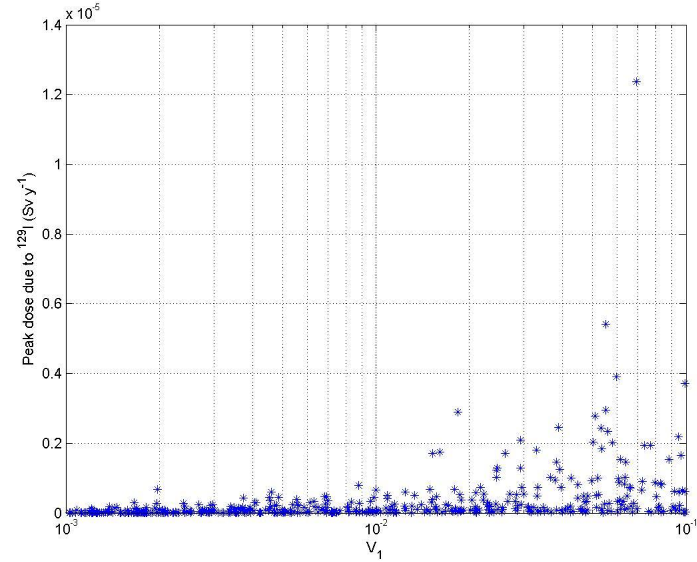
Methods:

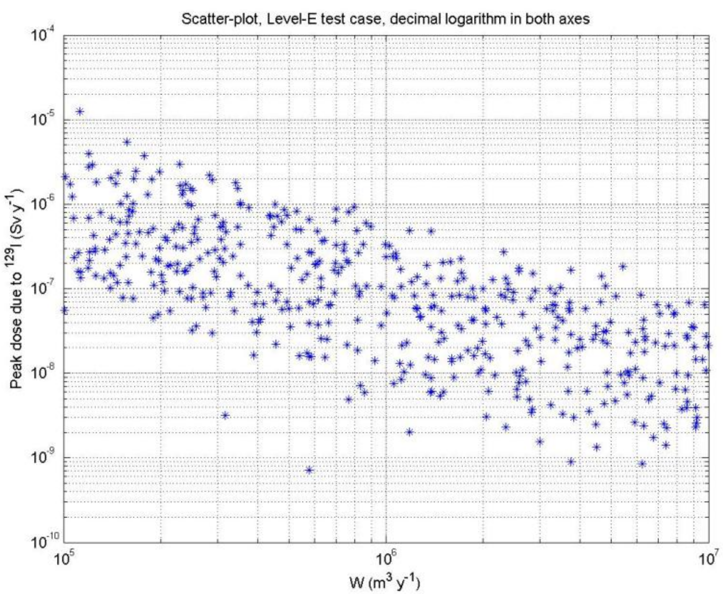
- Factorial Designs (FD)
- Fractional Factorial Designs (FFD)
- Andres' Iterated FFD (IFFD)
- Morris' one-at-a time designs
- Sequential bifurcation

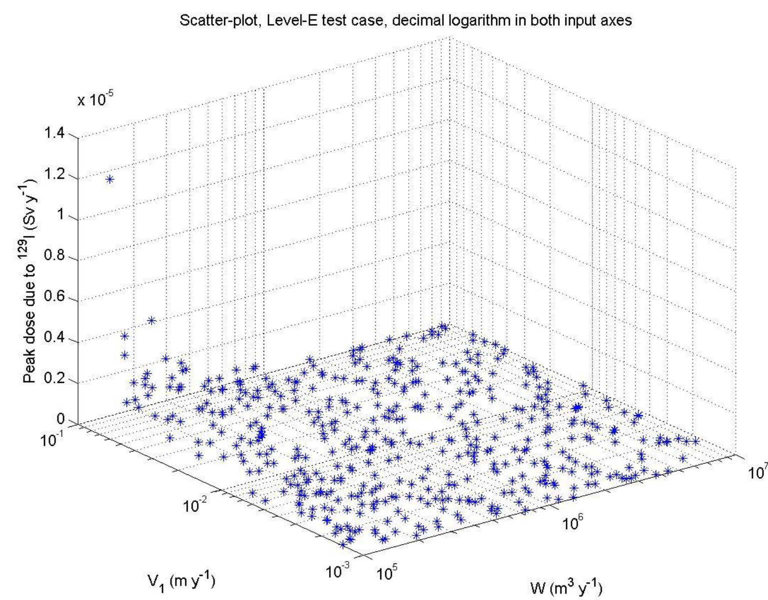


SCATTER-PLOTS

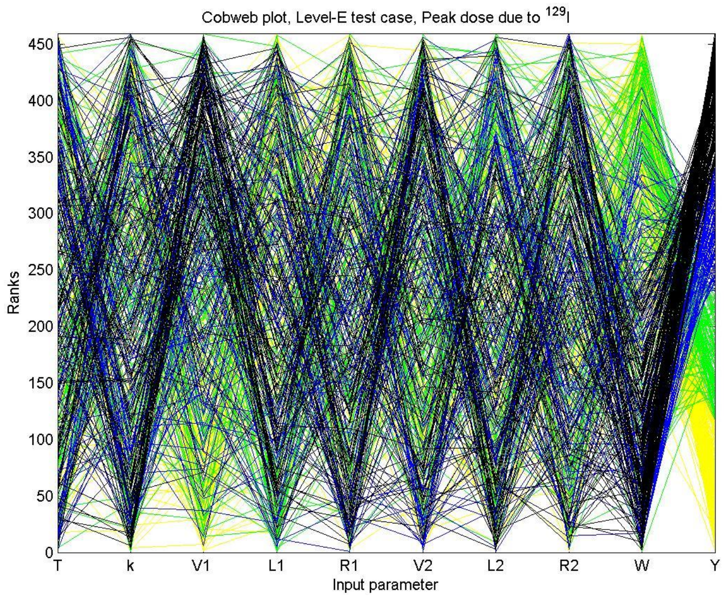


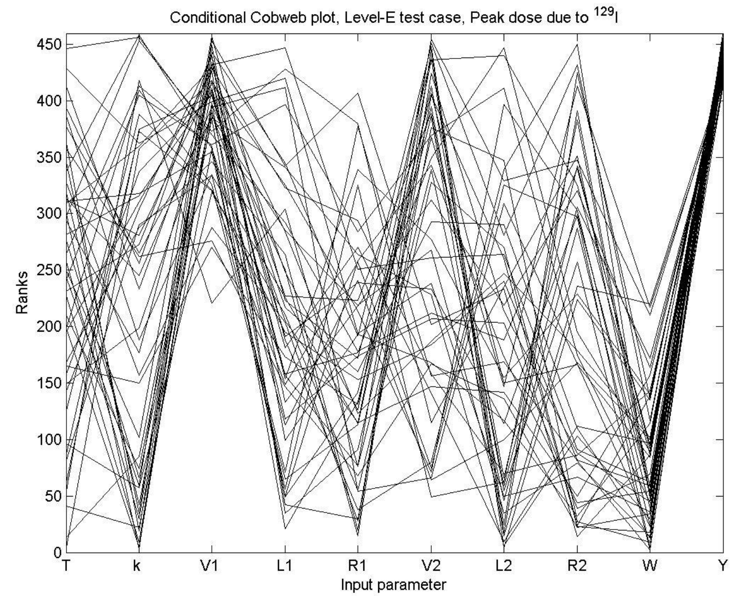


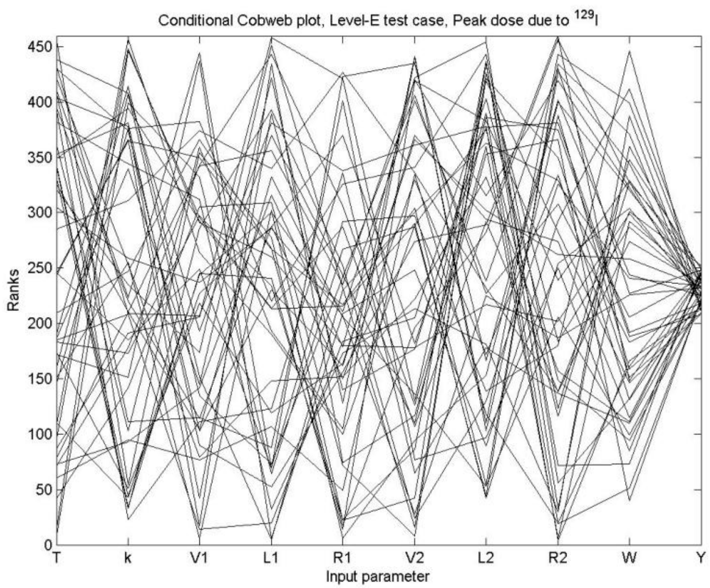


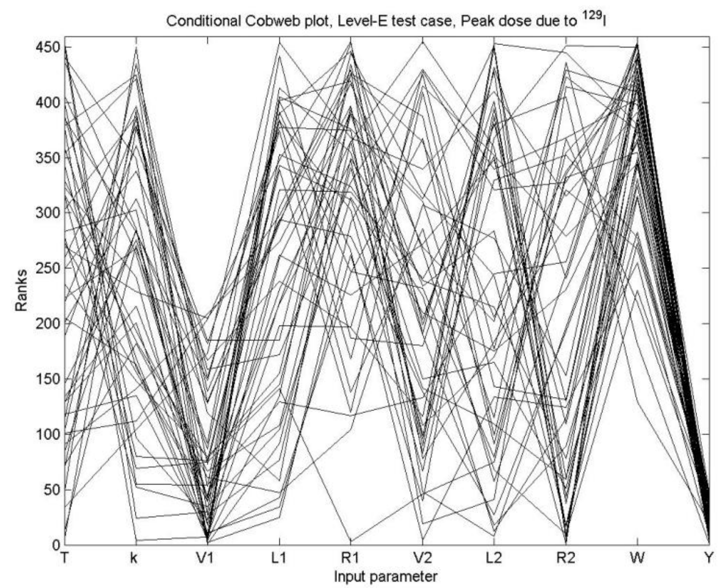


COBWEB-PLOTS



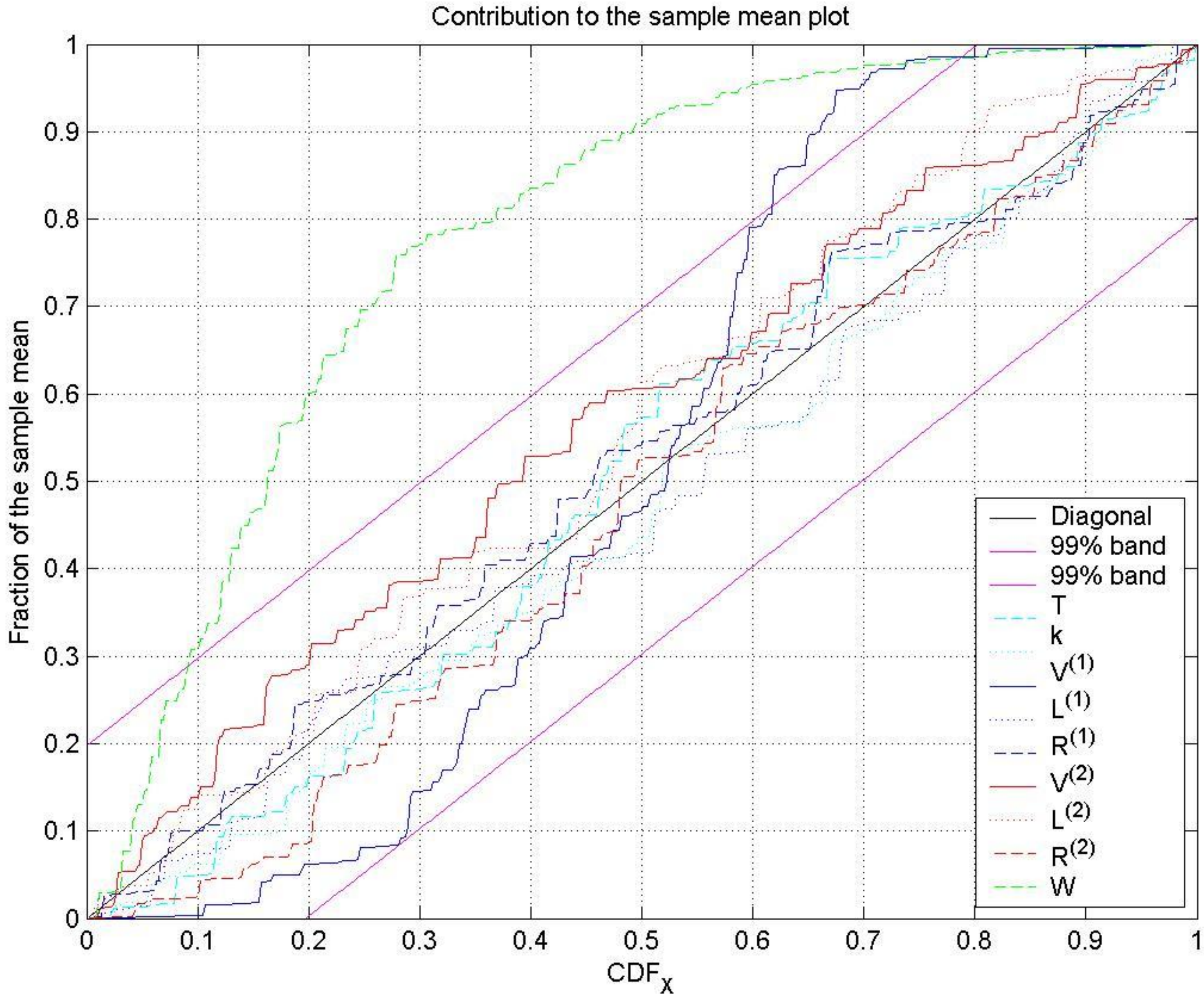






CONTRIBUTION TO THE SAMPLE MEAN PLOT – CSM plot

Statistical test developed under PAMINA



Regression based methods

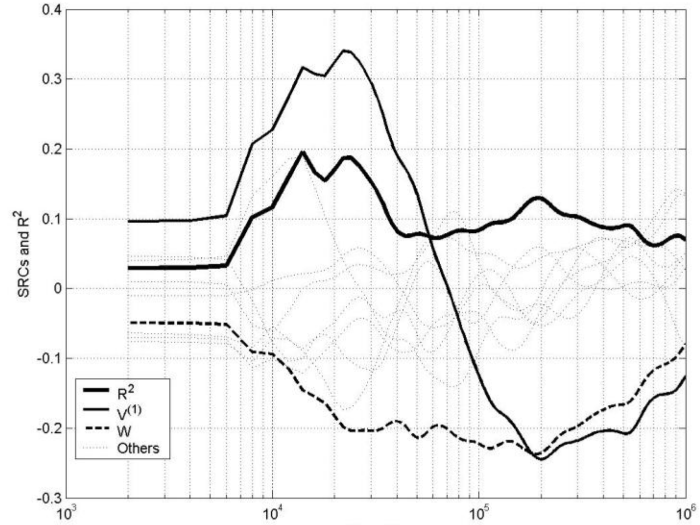
The PA model is studied assuming that it may be correctly represented by the following generic model (usually only up to main effects)

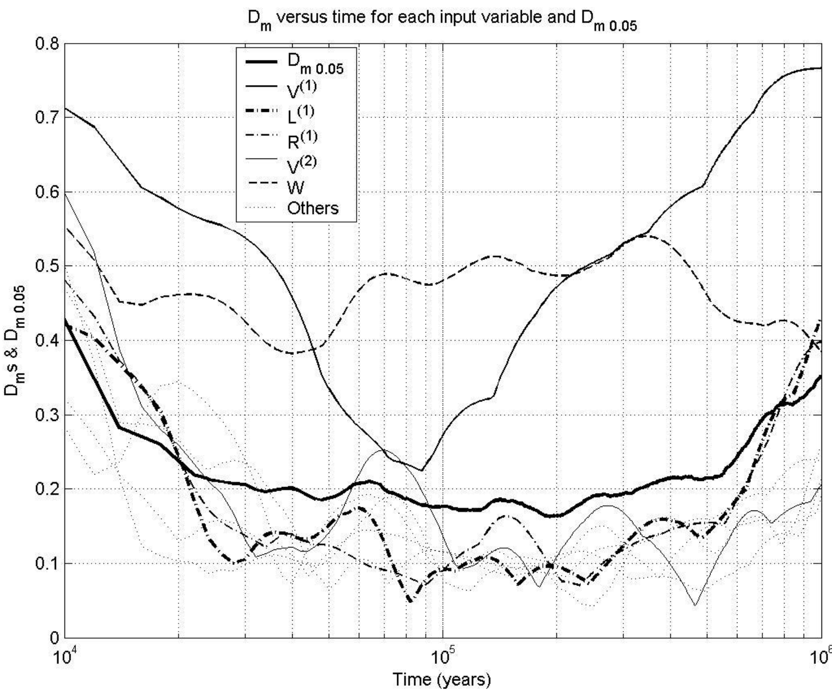
$$\hat{f}(\mathbf{x}) = \hat{\beta}_0 + \sum_{i=1}^{i=n} \hat{\beta}_i x_i + \sum_{i,j} \hat{\beta}_{ij} x_i x_j + \ldots$$

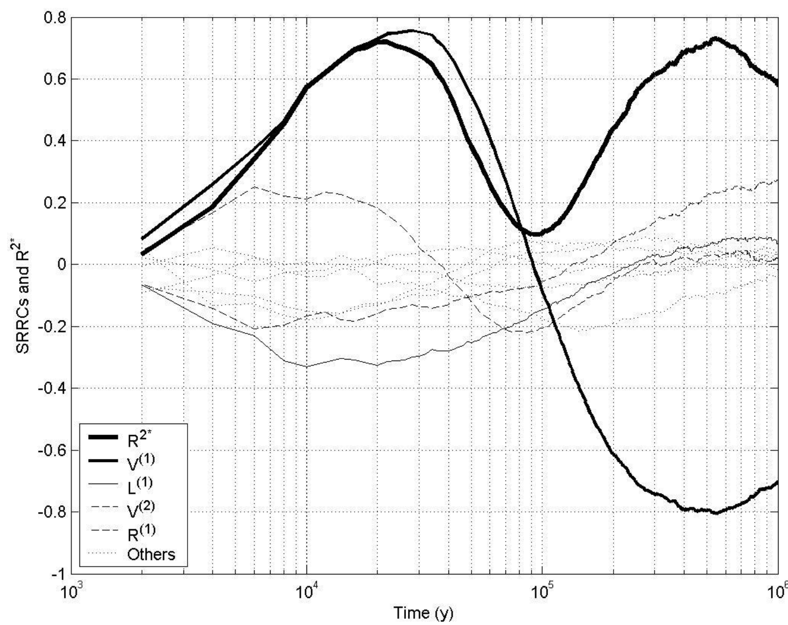
This model is estimated using standard statistical estimation techniques

- The output variable and the input parameters are standardized (subtract average and dividing by standard deviation)
    - Partial correlation coefficients (PCC)
    - Standardized regression coefficients (SRC)
  - R^2 is a measure of the quality of the results obtained
- Main problem: The assumed relation is too simple (no interactions, no higher order effects). Possible solution ➔ transformation of inputs and output (log, ranks, etc.)
- If ranks are used ➔ PRCC's & SRRC's

Regression based methods







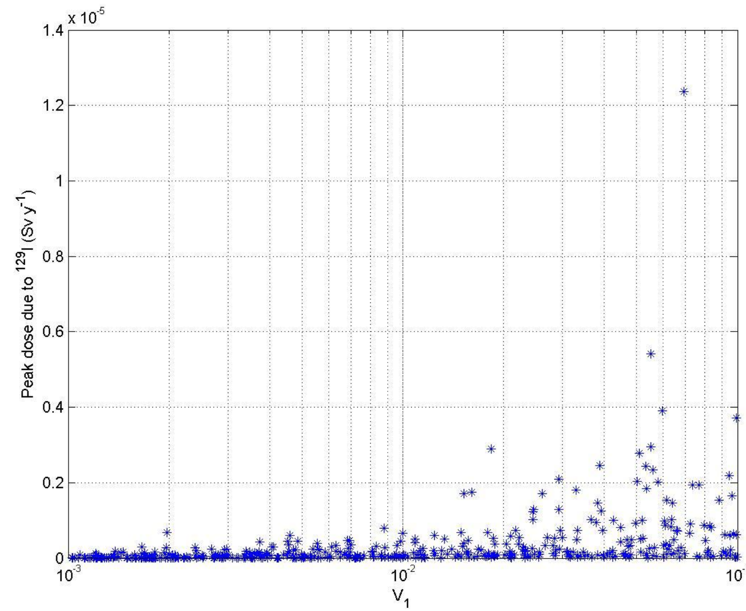
Monte Carlo filtering

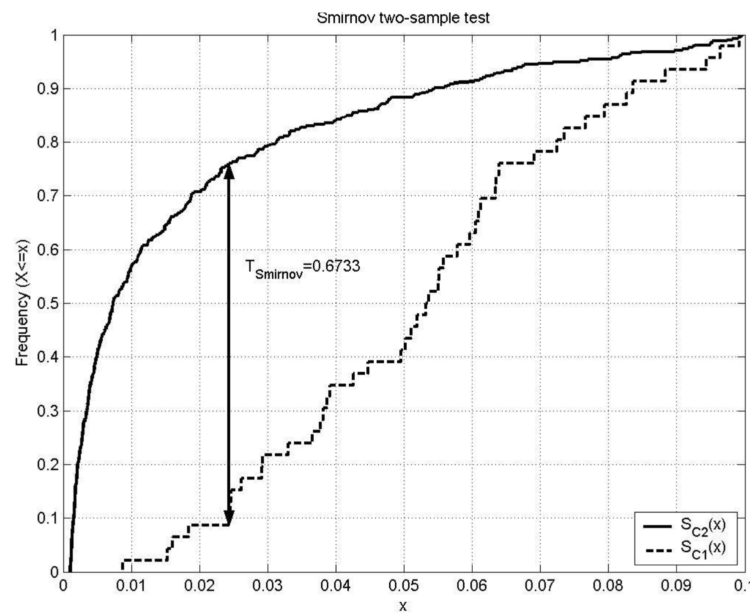
In some models there are neither clear linear nor monotonic relations, but there are threshold effects, discontinuities, …

In those cases, some classical statistics and graphic tools may be useful

- Mann-Whitney test,
- Kolmogorov-Smirnov test,
- Smirnov k-sample test
- Kruskal-Wallis test,
- Help of plots

p_value=0.0





Variance based techniques

The objective of these techniques is to ascertain what part of the output variable variance is due to each input parameter or interaction of parameters.

- Theoretical basis: Sobol's HDMR ➔ $D = \sum_{i=1}^{i=n} D_i + \sum_{i<j} D_{ij} + \ldots + D_{12\ldots n}$
- Sensitivity indices ➔ $S_{i_p \ldots i_h} = \dfrac{D_{i_p \ldots i_h}}{D}$
- Techniques:
  - Correlation ratios
    - Does not need specific sampling techniques. Only main effects estimated
  - Sobol's sensitivity indices
    - Very powerful, but needs specific sampling and huge sample sizes
  - Fourier Amplitude Sensitivity Test (FAST)
    - Powerful, but needs specific sampling and large sample sizes
    - Difficulties to properly estimate sensitivity indices for discrete input parameters (GRS- Braunschweig)
    - TU Clausthal is developing a new method to estimate SA indices using random samples

Sensitivity Analysis (SA) activities under PAMINA (RTDC's 2 & 4)

1. Review of SA techniques
2. Benchmark of SA techniques (8 partners participating)
   - Step 1: Study of 12 mathematical models
     - Test and debug SA tools
     - Starting from rather simple models, introduce complexity progressively
       - increase the number of parameters
       - add non-linearities
       - consider non-monotonic models
       - include periodicity
       - consider continuous models whose derivative does not exist at some given points
       - consider models with interactions
       - check the different capability to estimate accurately large and small sensitivity indices
     - Study the effect of sample size on the convergence and accuracy of estimates

Sensitivity Analysis (SA) activities under PAMINA (RTDC's 2 & 4)

– **Step 2**: Application of SA techniques to a simplified PA model (33 parameters, 12 of them uncertain). 4 radionuclides (^{129}I and the decay chain ^{237}Np, ^{233}U and ^{229}Th). Strong interactions.

3. Development of SA studies for different PA models proposed by partners (7 partners)
   – Different number of input parameters
   – Different computational cost per run
   – Different SA techniques

Several papers expected to be sent to scientific peer reviewed journals

Partner	PA case	PA model/ computational cost	N. of input parameters	SA methods	SA software
GRS-B/ JRC-Petten	Rock salt dome	EMOS-LOPOS TSS 30 min/run	6 All inputs independent	Regression based (linear & rank based) FAST/EFAST Smirnov	SimLab (JRC-Ispra) JRC-Petten software
	Indurated clay	EMOS-CLAYPOS TSS 30 min/run	<10 All inputs independent	Regression based (linear & rank based) FAST/EFAST Smirnov	SimLab (JRC-Ispra) JRC-Petten software
GRS-K	Iron ore mine	NAMMU-2D. Geosphere hydrology (saturated conditions)	~30 Dependences	Regression based (Spearman)	SUSA (GRS)
ENRESA/ JRC-Petten	Granite	GoldSim TSS 500-1000 run/day	~100 Independent	Some screening Most global	JRC-Petten software
	Plastic clay	GoldSim TSS 500-1000 run/day	~100 Independent	Some screening Most global	JRC-Petten software
NRG/ JRC-Petten	Salt	EMOS-ECN A few seconds/run	6 Independent	Most global	JRC-Petten software
	Soft clay	EMOS-ECN A few seconds/run	6-8 Independent	Most global	JRC-Petten software
Facilia	Granite	KBS-3 (focused on biosphere)	~100	All	EIKOS
ANDRA/ JRC-Petten	Indurated clay	Alliance platform, 2-D	~40 Dependences	Regression based (linear & rank based)	Alliance platform
	Indurated clay	Alliance platform, 3-D	~40 Dependences	Regression based (linear & rank based) MCF (Smirnov,...)	Alliance platform

Conclusions

1. Different sensitivity analysis techniques are available. Each one provides information about the input – output relations under a specific interpretation of the concept 'sensitivity'
2. The results provided by different techniques are complementary.
3. In general, most powerful techniques demand larger (in some cases unaffordable) sample sizes. ➔ a lot of room for improvement.
4. Relevant achievements within PAMINA
   - Deeper understanding of SA techniques
   - Real research in SA methods improving known methods
   - High degree of group integration, applying SA techniques to real PA model of interest to partners