Uncertainty and sensitivity methods in support of PSA level 2

Nicolas Devictor[*]
Ricardo Bolado Lavín[**]

[*]*Commissariat à l'Energie Atomique (CEA), Direction de l'Energie Nucléaire (DEN), Département d'Etudes des Reacteurs (DER), Service d'Etudes des Systèmes Innovants (SESI),*

13108 St-Paul-lez-Durance Cedex, France.

E-mail: nicolas.devictor@cea.fr

[**] *Nuclear safety Unit, Institute for energy – Joint Research Centre, European Commission*

P.O. Box 2, NL-1755 ZG Petten, The Netherlands.

E-mail: ricardo.bolado-lavin@jrc.nl

Abstract

Dealing with uncertainties in PSA level 2 requires using a set of statistical techniques to assess input uncertainty, to propagate uncertainties in an efficient way, to characterize appropriately output uncertainty and to get information from computer code runs through an intelligent use of sensitivity analysis techniques. The purpose of this paper is to give an overview of statistical and probabilistic methods and tools to answer to these topics, and to provide some guidance about their suitability and limitations to be used in a PSA level 2. Our position about their implementation in L2 PSA software has been written; it could be noticed that a lot of these methods are very time-consuming, and seem more suitable for the analysis of submodels or for focusing on specific questions.

1.- Introduction

PSA level 2 is a systematic way to study, from the point of view of safety and with the restrictions of a specific methodology, the behaviour of a system (NPP under accident or quasi-accident conditions) when uncertainty is present and widespread. The starting point of level 2 is the result of a PSA level 1. The results of such study is a huge quantity of accident sequences that are grouped, according to different criteria regarding accident characteristics and potential containment responses, into a manageable number of plant damage states (PDS). After an appropriate screening of very low probability sequences, the probabilistic progression of accidents is studied using event trees, commonly known as accident progression event trees (APET) or containment event trees (CET), under two possibilities: large event trees (virtually all questions regarding severe accident are included as top events) and small event trees (only main questions regarding severe accident phenomena are included as top events). The use of these event trees leads to getting a huge quantity of end states, that have to be grouped, as in the case of PDS's, to get a more manageable set of release categories, later used to estimate all the variety of different possible source terms. The appropriate combination of release categories and corresponding frequencies allows estimating the risk associated to the NPP.

Uncertainty is really pervasive in a PSA level 2. The first matter of concern is the starting point. Is really complete the picture obtained as an output from the PSA level 1? Could there be any 'hole' in the picture? Is completely meaningful the set of PDS obtained? In order to get the end states, computer codes that simulate in a deterministic way the behaviour of the plant have to be run under the containment event tree structure. Do they reproduce the behaviour of the system in an accurate way? Are we able to estimate accurately branching probabilities? Do we have the necessary data to feed the computer codes we are using? These and other questions are key issues when dealing with uncertainties in PSA level 2.

A lot of methods and tools do exist to study the influence of uncertainties on the results of severe accidents computer codes in use for PSA level 2. There are a lot of reports on the subject in the scientific literature [NEA 97, NEA 94, NEA 99, Devictor 04]. The item "influence of uncertainties" means, in this paper, uncertainty analysis or sensitivity analysis or evaluation of the probability of exceeding a threshold. These methods are sometimes not suitable, from a theoretical point of view, when the phenomena that are modelled by the computer code are discontinuous in the variation range of influent parameters, or some input variables are statistically dependent.

The purpose of this paper is to give an overview of statistical and probabilistic methods and tools to study the influence of uncertainties in input variables on the code responses, and to provide some guidance about their suitability and limitations to be used in a PSA level 2. In section 2 we will show different sources of uncertainty in PSA level 2, and different strategies to tackle them conditional on the quantity of available information. In section 3 we tackle the issue of uncertainty propagation, including the estimation of probabilities of exceeding a threshold and the use of surrogate models. In section 4 we describe main sensitivity analysis techniques, their usefulness and their meaning. Finally we show the main conclusions of this work.

2.- Uncertainty sources

Summarising the second paragraph of the introduction, we could say that uncertainty arises in three areas of the PSA level 2: 1) Definition of plant damage states, 2) Simulation of the problem, including event tree construction and models (computer codes) used to simulate the physical-chemical processes involved, and 3) data used to feed models. This is what classically has been considered scenario, model and data uncertainty. Nevertheless, attending to the real origin of uncertainty, these types of uncertainty may be re-classified as aleatory and epistemic or lack of knowledge uncertainty

2.1.- Aleatory and epistemic uncertainties

Aleatory uncertainties are usually associated to (chemical or physical) parameters with some inherent variability. Aleatory uncertainties arise when an experiment is repeated several times under equivalent conditions and the results obtained differ from each other. An example of a parameter affected by this kind of uncertainty is the time during which a battery provides the adequate electric current. In this case, variability comes from the set of physical and chemical processes involved. Increasing the number of observations (experiments) doesn't make aleatory uncertainty to decrease, but will allow us to know with more accuracy the probability density function (pdf) followed by that parameter, i.e. the type of pdf and the parameters that characterise it. So, if that time follows a Weibull distribution, increasing the number of observations will allow us to know more accurately the minimum failure time, the standard deviation and the shape parameter.

Epistemic uncertainties are related to the existence of lack of knowledge about the problem. This type of uncertainty affects not only parameters, but also models and PDS. A parameter will be affected by epistemic uncertainty when it isn't random, but we cannot measure it, either because it is impossible or because it is extremely expensive to do it. This type of uncertainty is completely different of the aleatory uncertainty. Parameters affected by aleatory uncertainty are fully described by their associated pdfs. In the case of parameters affected by epistemic uncertainty what we do is to characterise our uncertainty about the parameter, and we do it through pdfs. Those pdfs summarise our opinions about what values the actual value of the parameter could be close to more likely or less likely. Many parameters (coefficients) of models used in the area of severe accidents are affected by lack of knowledge uncertainty; they are not random, but we are unable to know their values, so we have to use pdfs to characterise them.

Epistemic uncertainty affects models. Sometimes, there are several models to describe the behaviour of the system; some of them describe the behaviour of the system under some circumstances and others under other circumstances, and it is not clear at all how to take that fact into account in the analysis. Think just of two different nodalisations for the same computer code. Some authors consider appropriate to assign probabilities to the different competing models and to run one of them or another one according to those probabilities. Other authors consider the right solution to build up a meta model that includes, as sub-models, the different models and runs either one sub-model or another one depending on the values sampled and what models fit better experimental results under those circumstances. Under any circumstance, only validated codes, or at least non-invalidated, should be used.

2.2.- Input uncertainty assessment

Input uncertainty assessment is the process of characterising through probability density functions (pdfs) or probability mass functions (pmfs) the uncertainty about continuous and discrete random variables. There are essentially two ways to do it: Using classical inference methods and using Bayesian methods. Expert judgment is a third way to do it, which constitutes, in this context, an extension of Bayesian methods.

2.2.1.- Classical inference methods

Classical inference methods are based on the assumptions of having a random sample and knowing the probability model from which the data come from. There are several methods, some of them recently developed, like Jackknife and Bootstrap methods, but the best known and most widely used methods are Maximum Likelihood method and the Method of Moments. The main shortcoming of all these methods is the need of sample sizes not easily obtainable under the restrictions of a complicated engineering facility in order to get good quality estimates.

Method of moments

This is probably the oldest inferential method to estimate the parameters of a pdf. K. Pearson developed the method of moments by the end of XIX century. The idea is quite simple. It consists in taking as an estimator of a parameter its equivalent sample quantity. So, the sample mean is the estimator for the mean, the sample variance is the estimator for the variance and so on.

Maximum Likelihood method

Maximum Likelihood Method is the most widely known and most powerful estimation method in the classical context. Let us assume that we wish to study a random variable X (parameter af-

fected by uncertainty) whose type of distribution function *f(X/θ)* is known, but whose parameter θ is unknown. In order to estimate θ we take a random sample - $\vec{X} = (X_1, X_2, ..., X_n)$, which is supposed to be a random vector whose components are independent and identically distributed (iid), so that its joint probability density function is

$$f(\vec{X}/\theta) = f(X_1, ..., X_n/\theta) = \prod_{i=1}^{n} f(X_i/\theta). \quad (1)$$

It is important to notice that in this expression, under the classical view, before sampling, θ is unknown, but has a given value, which determines what regions of $\vec{X}$ are more likely and what regions are less likely. So, this is a function whose unknowns are $\vec{X}$. This is the meaning before sampling. As soon as the sample is available, $\vec{X}$ is known, while θ remains unknown. The objective is to determine what value, among the infinite values that θ could take, makes more likely obtaining the sample actually obtained. So, the problem is to find the value of θ for which function (1) gets its maximum value.

2.2.2.- Bayesian inference methods

Bayesian interpretation of probability makes Bayes' formula a powerful tool to update degrees of belief when new information is available about an event or a proposition. Let *H* be the knowledge of a person, and let $\{z_i\}_{i \in I}$ be a partition of the sample space of events. The Bayesian probability provided by the person for an event z_k is $P(z_k/H)$. The acquisition of a set of new evidence *H'* induces a change in the probability given by Bayes' formula

$$P(z_k/H, H') = \frac{P(H'/H, z_k) \cdot P(z_k/H)}{P(H'/H)}, \quad (2)$$

where $P(z_k/H, H')$ is the 'a posteriori' probability of z_k, $P(z_k/H)$ is the 'a priori' probability of z_k and $P(H'/H, z_k)$ is the likelihood of evidence conditional on knowledge H and the occurrence of event z_k. $P(H'/H)$ is the probability of new evidence conditional on previous knowledge. Which means that the a posteriori probability is proportional to the a priori probability and to the likelihood of evidence.

Two remarkable results are obtained from (2). If the a priori probability of an event is zero, the a posteriori probability will remain zero, even though the evidence against it could be very strong. So, much care should be taken in order not to provide a priori probabilities. Null a priori probabilities should be avoided, unless total evidence of the impossibility of the events or propositions under study is available. The second result is related to the existence of strong evidence. In that case, likelihood will be completely dominant and the a priori probability will be almost irrelevant (a posteriori probability and likelihood will be almost equal). This is the case of large sample sizes, for which relative frequencies and Bayesian probabilities will be almost equal.

2.2.3.- Expert judgment

The use of Expert Judgment (EJ) techniques is unavoidable in a PSA level 2 due to the lack of data about many of the involved phenomena. In some cases, it is almost impossible, from a physical point of view, to get the data we need to feed our computer codes, in other cases the cost of getting them is so high that only a few of them may be obtained.In what follows, we will list the steps of a generic EJ protocol based on most widely known protocols available in scientific literature (see for example NUREG-1150):

1. Selection of team project
2. Definition of the questions to be studied
3. Selection of experts
4. Training
5. Tasks definition
6. Individual experts' work
7. Elicitation of experts' opinions
8. Analysis and aggregation of results
9. Documentation

Comprehensive information may be obtained in several references [Mengolini 05].

2.3.- What could be useful?

Whether a deterministic or probabilistic approach is implemented, sample or database treatment must be performed. The quality of the result of this analyze has a strong influence about the uncertainty results. A deterministic approach involves the identification of information like minimum, maximum values, envelope curves, etc., while a probabilistic vision concentrates on the dispersion or variability of the value, through variation interval or fractile-type data, or a probability distribution. Remember that a fractile or quantile of the order α% is a real number, $X_{\alpha,}$ satisfying $P(X \leq X_{\alpha}) = \alpha\%$. Treatment is compatible with the intended application as, for example, determining a good distribution representation around a central value or correctly modelling behaviour in a distribution tail, etc.

Tools to describe sample dispersion are taken from Statistics; however, their effectiveness is a function of the sample size. References [Saporta 90] and [Lannoy 01] describe methods that may be used to adjust a probability distribution on a sample, and then verify the adequacy of this adjusted distribution in the maximum failure probability region. It is obvious that if data is lacking or scarce, these tools are difficult to use. Under such circumstances, it is entirely reasonable to refer to expert opinion in order to model uncertainty associated with a value, and then transcribe said information in the form of a probability distribution.

The practical approach can be summarised by three scenarios:

Scenario 1. If a lot of experience feedback data is available, the frequential statistics is generally used. The objectivist or frequential interpretation associates the probability with the observed frequency of an event. In this interpretation, the confidence interval of a parameter, p, has the property that the actual value of p is within the interval with a confidence level α; this confidence interval is calculated based on measurements.

Scenario 2. If data is not as abundant, expert opinion may be used to obtain modelling hypotheses. The Bayesian analysis is used to correct a priori values established based on expert opinion as a function of observed events. The subjectivist (or Bayesian) interpretation understands probability as a degree of belief in a hypothesis. In this interpretation, the confidence interval is based on a probability distribution representing the analyst's degree of confidence in the possible values of the parameter and reflecting his/her knowledge of the parameter.

Scenario 3. If no data is available on a parameter, its probabilistic representation may be obtained from a model and from the knowledge of the uncertainties on the input parameter of this model. The data to be gathered thus concerns the input parameters. The quality of the probabilistic analysis is a function of the credibility of sta-

tistics concerning these input parameters and that of the model. The following cases can be discerned:

- A structural reliability-type approach if the sought value is a probability,
- An uncertainty propagation-type approach if a statistic around the most probable value is considered.

As shown in different studies, none approach is a panacea for the uncertainty modeling (see for example [Helton 04] and [WEJ 05]). At the present time, works for building a coherent mathematical theory for uncertainty that involves different paradigm are on-going (see for example the Demspter-Shafer theory or theory of evidence); these works seems not at the present time mature.

"Scenario 1", where a large enough sample is available, i.e. the sample allows "characterisation of the relevant distribution with a known and adequate precision", begs the following questions:

Question 1 :Is the selected distribution type relevant and justifiable? From the various statistical models available, what would be the optimal distribution choice?

Question 2 :Would altering the distribution (all other things being equal) entail a significant difference in the results of the application?

Question 3 :How can uncertainty associated with sample representativeness be taken into consideration (sample size, quality, etc)?

Justification is difficult for Scenarios 2 and 3. For example, if the parameters of a density are obtained from the first moments, it must be borne in mind that a precise estimation of the symmetry coefficient requires at least 50 values while kurtosis requires 100 data, except for very specific circumstances. Furthermore, the critical values used by tests to reject or accept a hypothesis are frequently taken from results that are asymptomatic in the sense that sample size tends towards infinity. Thus when sample size is small, the results of conventional tests should be handled with caution.

With respect to question 2, a study examining sensitivity to the used probability distribution provides information. There are two methods available for the sensitivity study:

- Moment identification-type method where it is assumed that the distribution changes while the first moments are preserved (mean, standard-deviation especially).
- Frequential or Bayesian approach where the sample is redistributed to establish the reliability data distribution parameters.

Another method is to take the uncertainty associated with some distribution parameters into consideration by replacing the parameters' deterministic value by a random variable. The thesis [Pendola 00] provides other information in response to questions 2 and 3.

Conventional criticism concerning the statistical modelling of a database concerns:

- Difficulty in interpreting experience feedback for a specific application;
- Database quality, especially if few points are available;
- Substantiation of the built probabilistic model.

The probabilistic modelling procedure should attempt answer these questions. If it is not possible to define a correct probabilistic model, it is obvious that, under these circumstances, the quantitative results in absolute value are senseless in the decision process. However, the probabilistic approach always allows results to be used relatively, notably through:

- a comparison of the efficiency of various solutions from the standpoint of reliability, availability, for example,

- or classification of parameters that make the biggest contribution to the uncertainty associated with the response in order to guide R&D works in order to reduce said uncertainty.

The deterministic approach involves validation of values used and also constitutes a sophisticated problem: it is not easy to prove that a value assumed to be conservative is realistic, especially if the sample is small or is a guarantee that the value is an absolute bound must be provided. Frequently, conservative values used are formally associated with small or large fractiles of the orders of the values studied. The concept of fractile is associated with the probability distribution adjusted on the sample, and even with one of the distribution tails.

The probabilistic approach seems to be even more suitable to deal with the problem. In fact, the probabilistic model reflects the level of knowledge of variables and models, and confidence in said knowledge. By means of sensitivity studies, this approach allows the impact of the probabilistic model choice on risk to be objectively assessed. Furthermore, in the event of new information impugning probabilistic modelling, and consequently the fractiles of a variable, the Bayesian theory, that combines objective and subjective (expertise) data, allows the probabilistic model and results of the probabilistic approach to be updated stringently.

The adjustment of a probability distribution and subsequent testing of the quality of said adjustment around the central section (or maximum failure probability region) of the distribution constitute operations that are relatively simple to implement using the available statistical software packages (SAS, Stagraphics, Splus, Statistica, etc.), for conventional laws in any case. However, the interpretation and verification of results still requires the expertise of a statistician. For example, the following points:

- the results of an adjustment based on a histogram is sensitive to the intervals width;
- the maximum likelihood or moment methods are not suitable for modelling a sample obtained by overlaying phenomena beyond a given limit of an observation variable;
- moment methods assume estimations of kurtosis and symmetry coefficients that are only usually specified for large bases (with at least one hundred values for kurtosis);
- some statistical tests are based on asymptotic results;
- in the Bayesian approach, the distribution selected *a priori* influences the result. Furthermore, the debate concerning whether or not the least informative law should be used has not been concluded.

In industrial studies, the characteristics of databases must be taken into consideration because the implementation of an adjustment can be difficult; for example the following frequently encountered scenarios:

1. the sample size is small and therefore asymptotic results need to be handled cautiously as well as approximations of more or less valid moments of an order greater than two;
2. sample data values are measured with an uncertainty;
3. sample homogeneity is not verified (mixture of samples taken from different populations, overlaying of phenomena, etc.);
4. if the area of interest is a distribution tail, it should be noted that the statistical theory and above all associated tools are less developed.

2.3.1.- Selection of a distribution in practice

In practice, the criteria relevant to selecting a probabilistic model for a random variable would seem to be:

- use a family compatible with the physical properties (bound value, symmetrical or not, exhibiting exponential decay, etc),
- the result of data adjustment,
- distribution not rejected by statistical tests,
- the selection of a distribution that is least "informative" with respect to data or available information (i.e. the introduction of too many hypotheses is avoided).

Statistical tests are used to decide whether or not the adequacy of a selected distribution should be rejected a priori with a confidence level. For samples that are not very homogeneous or small in size, several distributions are frequently accepted to represent the sample or, on the contrary, no conventional distribution can be accepted.

2.3.2 Some criticism

To select a specific distribution, the following methods can be applied:

- either rely on selection based on expert opinion or current practice like, for example, a Weibull distribution for reactor vessel steel toughness measurements ([Wallin 84]) or a logarithmic normal distribution for a constraint;
- or compare confidence levels taken from various statistical tests for each accepted distribution and select the distribution associated with the highest confidence level;
- or further analyze adequacy in the zone of interest and select the distribution most graphically suited in the region of interest (for example a distribution tail for a reliability problem).
-

Another possibility is to use a resampling method: for example the Bootstrap method or cross-validation method, in order to select "the most relevant" distribution family (ref. for example [Chevallier 97] and [Devictor 99]).

As regards the non-parametric Bootstrap method, "the most relevant" family is defined as that which is almost always accepted by adjustment tests. This method assumes that sample size is not too small. Experience has also shown that the non-parametric Boostrap method is rarely suited to the treatment of extreme values.

3.- Propagation of uncertainties

In a PSA level 2 it is necessary to estimate as accurately as possible all relevant output variables. The full characterisation of a random variable is given by its probability density function or, equivalently, by its integral, the cumulative distribution function (CDF), or its complementary curve to 1 (CCDF). A part of the information contained in those curves is summarised by some numeric statistics like the mean, the standard deviation and order statistics, among others. Additionally, there are several graphics that provide visual information about the shapes of the aforementioned functions.

3.1.-Monte Carlo method

The Monte Carlo method is based on sampling the vector of input parameters, running the system model computer code for each sample of that vector in order to get a sample of the vector of output variables, and estimating the characteristics of the output variables using the output samples obtained. One of the benefits of using the Monte Carlo method is that we may use all statistical standard estimation methods and tests to estimate the output variables distributions and to test any

hypothesis. This makes it the most straightforward and powerful method available in the scientific literature to deal with uncertainty propagation in complex models.

3.1.1 Sample quality check

After sampling, and before running the computer code, the quality of the sample obtained for the vector of input parameters should be checked. This means that the sample fits the multidimensional probability density function assigned to the vector of input parameters. This task is almost impossible to be done if the structure of that random vector is very complicated (complicated dependence structure between vector components) due to lack of adequate tests. Nevertheless, in many practical cases, most of the input parameters are stochastically independent, or dependence is implemented as correlation coefficients in the case of Gaussian variables. In those common cases, the check of the sample quality may be done with a reduced effort. In fact two things have to be checked:

* Fit of sampled and theoretical distribution for each variable (input parameter)
* Correlation coefficients matrix

The first check consists of performing as many goodness of fit tests as input parameters are sampled. There are many different tests that could be adequate for this check, some of them are designed ad hoc for some specific types of pdfs, others are generic. Among the latter, the best-known tests are Kolmogorov's test and χ^2 test. Since Kolmogorov's test is not based on asymptotic properties of samples, this is the preferred one in the opinion of the authors. So, if there are k parameters, k tests should be done. When performing this set of k tests there are two possible reasons to reject the sample: 1) The number of variables wrongly sampled is large and 2) one or more variables produce very bad results of the goodness of fit test (it is not yet defined what is a very bad result, but as a guide a bad result could be an individual goodness of fit test with a p-value smaller that 10^{-3} or 10^{-4}).

Checking the correlations between different parameters is a little bit more complicated because there could be many different cases. In this text let us assume the simplest case: all variables are independent. Under these circumstances, the sample correlation coefficient between each couple of variables should be close to 0 (the theoretical one is 0). The adequate test to check if those sample correlation coefficients are too large is the test of Pearson. As in the case of the goodness of fit tests, there are two possible reasons to reject the sample: 1) The number of large correlation coefficients is too large and 2) there are one or more correlation coefficients that are too large. Again, it is not clear what a large correlation coefficient is, but correlation coefficients that produce p-values smaller that 10^{-3} or 10^{-4} could be considered too large.

3.1.2.- Most commonly used statistics

The most widely used statistics to summarise the information contained in a sample are the sample mean ($\hat{x}$) and the sample standard deviation ($\hat{S}$). They are respectively estimators of the mean and the standard deviation of the variable. Additionally, both of them are used to provide approximate confidence interval for the mean of the studied variable.

Quite important statistics are the order statistics. Given a sample of size n $X = \{x_1, x_2, ..., n_n\}$, the order statistics obtained from this sample are $x_{(1)}, x_{(2)}, ..., x_{(n)}$, where the number between brackets stands for the number of order ($x_{(1)}$ stands for the smallest observation in the sample, $x_{(n)}$ stands for the largest observation). There are two important facts related to these statistics: 1) Any order

statistic $x_{(i)}$ is an asymptotically unbiased estimator of the i/n quantile of the studied variable and 2) based on the their distributional properties, exact distribution-free confidence intervals may be provided for any quantile. Distribution-free means that the confidence intervals generated do not rely on assuming any specific type of distribution for the variable. This is an extremely important result since allows to provide estimates for any percentile (the median, 95%, 99%, etc.) of a random variable and exact confidence intervals (for large enough sample sizes).

3.1.3.- Uncertainty range and confidence interval estimation

Usually, Safety Goals are defined in terms of some specific statistical quantities, normally means or some pre-specified quantiles (50%, 75%, 90%, 95%, 99%, 99.9%, 99.99% , etc.). Nevertheless, providing just the estimates of those quantities doesn't say much about the quantities themselves. One observation is enough to provide an estimate of the mean and two observations are enough to provide an estimate of the median (quantile 50%) and the standard deviation, but, certainly, nobody would trust the estimates provided by samples as small as those mentioned in the last sentence. The question is: how large should our sample be in order to provide meaningful and believable estimates? The answer to this question is provided by the theory of interval estimation (confidence intervals) and by tolerance intervals results (based on order statistics theory), and depends on what we want to estimate and to guarantee.

Interval estimation is a part of statistical inference whose objective is to generate confidence intervals for different quantities. A confidence interval of a given quantity, like for example of a mean, a standard deviation, a probability or a median, is, roughly speaking, an interval that will be based on the sampled data, and, a priori, before computing it, we know that will contain, with a pre-specified probability, the quantity to be estimated. As previously mentioned, provided large enough samples, exact confidence intervals may be provided for any quantile. The case of the mean is completely different, exact confidence intervals are available only for a few distributions, such as normal or exponential distributions. For any other distribution, only approximate intervals are available, which are based on the asymptotic normality of the mean. Such property is not easily demonstrated.

A tolerance interval (L,U), usually determined by the maximum and the minimum observations in a sample, is defined as an interval for which we can guarantee that, with a pre-specified probability (β), a given proportion (γ) of the random variable is contained within them. In mathematical notation this is

$$\Pr\left(\int_L^U f_x(x)dx \geq \gamma\right) = \beta \quad . \tag{3}$$

In this sense we will speak about, for instance, a 95%/99% tolerance interval (γ=*0.*95, β=*0.99*), as an interval that will contain 95% of the values of the random variable under study with probability 99%. The problem to be solved is how large should our sample be to guarantee with probability (at least) β that a proportion (at least) γ of the studied variable is contained between the limits L and U to be estimated. 'At least' comes from the fact that is impossible to fit exactly those to conditions imposed, so that we try to be on the safe side. Tolerance intervals may be two-sided or one sided. A two sided interval is obtained when no restriction is imposed on U and L; a one-sided interval is obtained when either U or L are fixed (either to a physical threshold or to $\pm\infty$). The following table shows minimum sample sizes to get such tolerance intervals [Wilks 41].

Table 1.- Minimum sample size for one and two sided tolerance intervals for different values of β and γ.

	Two sided tolerance intervals			One-sided tolerance intervals		
B \ γ	0.90	0.95	0.99	0.90	0.95	0.99
0.90	38	77	388	22	45	230
0.95	46	93	473	29	59	299
0.99	64	130	662	44	90	459

3.1.4.- Estimating the probability of failure of a process

Using the Monte Carlo method to compute the probability of failure of process consist of applying the method as usual, using as an output variable a performance function. The value of this variable is compared for each run against a failure criterion. If the failure criterion is fulfilled, that run is considered a failure. The estimate of the probability of failure is

$$\hat{P}_f = N_f / N \tag{4}$$

where N_f is the number of runs in which the failure criterion is fulfilled, and N the total number of runs. As N approaches infinity, $\hat{P}_f$ approaches the true probability of failure. The accuracy of the estimation can be evaluated in terms of its variance computed approximately as:

$$Var(\hat{P}_f) \cong (1 - \hat{P}_f)\hat{P}_f / N \tag{5}$$

The smaller the variance, the better the accuracy of the estimated probability of failure. It is obvious that as N approaches infinity, $Var(\hat{P}_f)$ approach zero. For a small probability of failure and a small number of runs, the variance of $\hat{P}_f$ can be quite large. Consequently, it may take a large number of runs to achieve a specific accuracy. The amount of computer time needed for the direct Monte Carlo method is large, specially in the case where each run involves a long calculation performed by a thermal-hydraulic code. Variance reduction techniques could be of use to reduce computational costs.

3.1.5.- Variance reduction techniques

In many actual applications, applying simple Monte Carlo may be prohibitive, conditional on the code used. In those cases, variance reduction techniques could be of use to reduce the computational costs. Variance reduction techniques allow the user to get the same accuracy with a lower computational cost. Main techniques are Latin Hypercube sampling (LHS), stratified sampling, control variates, importance sampling and antithetic variates, among others.

3.1.6.- The use of surrogate models

Some of the computer codes used to simulate severe accidents or some Level 2 PSA models are very expensive in computational terms, so that, in many cases, is unthinkable to run them as many times as needed to get accurate probabilistic results. In those cases it could be of interest to seek for good predictors that could be used as surrogate models (or synonym “response surface” or “metamodel”) to be run instead of the real models.

Principles of the method (RSM)

The response surface method (RSM) [Box 87] is used to build a function which simulates the behavior of a physical or chemical phenomenon in the field of variation of the influential parameters, starting from a certain number of experiments. It was originally proposed as a statistical tool, to find the operating conditions of a chemical process at which some response were optimized. Subsequent generalizations led to these methods being used to develop approximating functions that surrogate for long running computer codes ([Sacks 89], [Devictor 00], [Iooss 05]). These surrogate response surface models fit data that are generated by running the simulation model at selected points in the parameter space. Building such response surface aims at obtaining a mathematical model representative of the studied software, having good capacities of prediction, and for which computing time to evaluate an output variable is negligible. It will be thus efficient for the uncertainty and sensitivity analyses requiring several thousands of simulations.

There are different issues involved in selecting runs for uncertainty analysis and correlation coefficients computation compared with selecting runs for building a response surface ([Sacks 89]). For this latter purpose, one can often do better than random selection by making hypotheses on the response surface model and on the interactions between parameters. There are multiple response surface families [Hastie 02]: polynomials, splines, interpolating radial functions, kriging, generalized linear models (GLM), partial least squares, neural networks, regression trees, support vector machines, etc.

In general, to build a response surface, it is necessary to have the following elements:

- a sample **D** of points (**x**(i), z(i)), where **x** is the vector (x_1, …, x_p), $z=f(\mathbf{x})$;
- P(**X**,Z) the probability law of the random vector (**X**,Z) (unknown in practice);
- a family F of response surface functions f(**x**,**c**), where **c** is either a parameter vector or a index vector that identifies the different elements of F.
-

The aim of a RSM is, in consideration of the sample **D**, to determine the best representing function f_0 in **F.** It is then the function f_0 that minimizes the risk function:

$$R(f) = \int L(z, f(\mathbf{x},c))dP(\mathbf{x},z) \tag{6}$$

If assumptions of normal distributions and constant standard deviations are done, the loss function is quadratic.

Choice of design points

In case of surrogate model of simple shape, building points or sample **D** are often chosen from classical design of experiments. In case of polynomial surrogate models, main designs used are 2^k, 3^k, star designs and other composite designs. According of the characteristics of the code, other experimental designs are available in the literature, as D-optimal designs, etc. During the last years other options have been explored, mainly seeking for designs that could detect discontinuities. As the use of surrogate models introduces bias errors, these models should be used only if the reference model is too much time consuming.

Before any use of response surface, it is necessary to qualify it for the foreseen utilisation. This qualification keeps a part of subjectivity. The characteristics « good approximation » is subjective and depends on the use of the response surface. The use could introduce additional constraints. For example, constraints like “conservatism”, “bound on the remainder”, “a better accuracy in an interest area (distribution tail…)” can be needed. For reliability studies, a good representation of the domain of maximum of failure probability is often sufficient and it is not necessary to seek a

good quality of approximation in the entire field of variation of the input parameters. If the response surface is used in a problem where the knowledge of uncertainties is inaccurate, it is not judicious to seek response surfaces explaining 99.9% of the variability.

Validation of a response surface

The statistical accuracy of the response surfaces can be assessed by cross-validation or bootstrap techniques ([Hastie 02]). The cross validation method allows a good estimation of the theoretical prediction error associated with the response surface, while the bootstrap method is especially useful when the size of the data sample is small. With a sufficient size of data sample, a simpler method can be chosen:

1. the quality of approximation is given by statistical analyses carried out on the bases of points used to build the surface (this set of points is called here "training set");
2. the quality of prediction is obtained by statistical analyses carried out on points not belonging to the building base (this set of points is called the "base of test").

The following general methodology can be used to validate the response surface:

- Initially some indicators obtained from the response surface with those obtained directly with software are compared on the two databases: average, standard deviation, minimum and maximum.
- A regression analysis allows determining the share of variability of the output variable explained by the fitted model. Two statistics, which give global measurements of correlation between two data sets, have been particularly studied: the Pearson correlation coefficient and the coefficient of determination R^2. These statistics between the software response and the response surface are calculated. The values of these statistics have to be approximately equal using the points of the database of construction and those of the database of prediction.
- The previous criteria are global and it is possible that the adjusted data are not homogeneous. This is the case when the studied variable covers a broad range of variations with multiple orders of magnitude. In this case, the contributions of the low values to the R^2 measurement are negligible. To cure this problem, the study of residual statistics gives some indications of the regression accuracy. The residuals $\Box(\mathbf{x},f)$ have to follow a Gaussian distribution of mean zero, with a constant standard deviation small compared to the standard deviation of Z (results of the computer code).
- The average and the standard deviation of the relative residuals $\Box(\mathbf{x},f)/z(x)$ are also examined. The response surface is valid on all the field of variation if the average and standard deviation of these relative residuals are small.

3.2.- Other methods

In the following text we describe two more methods to propagate uncertainties. The first one is a general method designed to estimate the distribution of an output variable, while the second one is designed to estimate a probability of failure.

3.2.1.- Method of moments

This method allows to calculate the first four moments by using the Gauss integration method and then to fit a distribution of probability to these moments by using the Pearson or Johnson methods (cf. [Pearson 65] and [Baldeweck 97]). The first objective is to evaluate the first moments of the random response *Y*. The expectation of *Y* can be calculated by:

$$E(Y) = E(g(X)) = \int \cdots \int g(x) f_X(x) dx_1 \cdots dx_p \tag{7}$$

where f_x is the joint density distribution of the input uncertain parameter vector X.
This equation can be evaluated by a squaring method of Gauss. This method allows the integration of a continuous function with the desired accuracy. It consists in the discretisation of the interval of integration in a number of *X*-coordinates x_i to which a weight w_i is associated. The number of *X*-coordinates is a function of the desired accuracy. For a continuous function g(x):

$$\int_a^b W(x) g(x) dx \approx \sum_{i=1}^{N} \omega_i g(x_i) \tag{8}$$

Practically, a set of order j orthogonal polynomials $\{p_j(x)\}_{j=0,1,2\ldots}$ are associated to the weight function W(x). These polynomials verify the following relations:

$$\int_a^b W(x) p_i(x) p_j(x) dx = 0 \qquad \text{if } i \neq j; \qquad \int_a^b W(x) p_i(x)^2 dx = 1 \tag{9}$$

The *N* X-coordinates of a squaring formula with a weight function *W(x)* are the zeros of the polynomial $p_N(x)$, which has exactly N zeros in the interval [a, b]. Relations of recurrence generally define these polynomials. The weights are calculated by solving the system of linear equations:

$$\sum_{j=1}^{N} p_i(x) \omega_j = \int_a^b W(x) p_i(x) \qquad i = 1 \cdots N \tag{10}$$

Then the average is evaluated from:

$$E(Y) = E(g(X)) \approx \sum_{i=1}^{N} \omega_i \times g(u_i) \tag{11}$$

and the moment of order *k* from:

$$M_Y(k) = \int [g(x) - \mu_Y]^k f_X(x) dx \approx \sum_{i=1}^{N} \omega_{i,X} \times [g(u_{i,X}) - \mu_Y]^k \tag{12}$$

Pearson and al. ([Pearson 65]) show that one can define in an approximate way a density of probability from the average, the standard deviation and two additional coefficients called coefficients of Fisher: The coefficient of symmetry and the coefficient of flatness.The four-parameter Pearson family of distributions can approximate a great number of continuous distributions. The curves can have several shapes (bell-shaped curve, curved in J, curved in U). This method is efficient to estimate a *pdf* if the number of random variables is small.

3.2.2.- Approximate methods (FORM/SORM)

The first- and second-order reliability methods (FORM/SORM, see [Melchers 99]) are suitable methods for computing probabilities of failure of a system. They consist of consist of 4 steps:

- the transformation of the space of the basic random variables $X_1, X_2,\ldots,X_n$ into a space of standard normal variables (Rosenblatt transformation when the joint density is known, and the transformation of Nataf when the probabilistic model is only made up of the marginal densities and of the matrix of covariance),

- the search, in this transformed space, of the point of minimum distance from the origin on the limit state surface, which is called 'design point'(The Hasofer-Lind index β_{HL} is defined as the minimal distance between the failure surface and the origin of the Gaussian space),
- an approximation of the failure surface near the design point (either linear – FORM, or quadratic - SORM),
- a computation of the failure probability corresponding to the approximating failure surface.

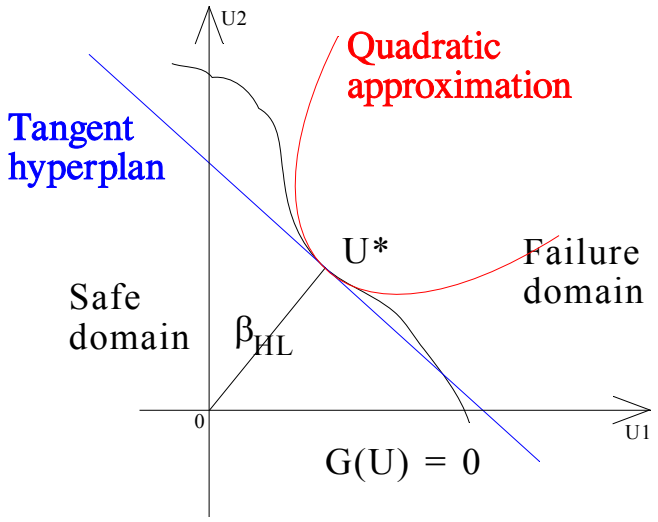

Figure 1: Reliability assessment with FORM/SORM methods

Characteristics of FORM/SORM methods

FORM and SORM are approximate methods, and their accuracy is generally good for small probabilities. The analytical properties enable the methods to yield relatively inexpensive sensitivity factors. The basic random variables must be continuous, and the failure function must be continuous. With the optimisation procedures presently used in most cases, the failure functions should be smooth.

Accuracy: The methods are approximate, but yield generally accurate results for practical purposes, in particular for small range probabilities (10^{-3}-10^{-8}).

Efficiency: For small order probabilities FORM/SORM are extremely efficient as compared to simulation methods. The CPU-time is for FORM approximately linear in the number of basic variables n, and the additional CPU-time for a SORM computation grows approximately with n^2. The absolute computation time depends on the time necessary to evaluate the failure function. This time may in fact depend on the actual values of the basic variables. Extreme values may take longer due to increased non-linearities in the problem. The CPU-time is independent of the probability level, assuming a constant time for the evaluation of the failure function.

Restrictions and further developments: When the failure surface is not sufficiently smooth, the most likely failure point cannot be identified by efficient mathematical programming methods applying the gradient of the function. In this case, it would be advantageous to fit the exact failure

function by a differentiable function. If the function is extremely CPU-time costly to evaluate, a simpler function may be used. It was suggested to use response surface methods. In particular, linear and quadratic response surfaces are well suited in combination with FORM/SORM, since the exact result with respect to the response surface is extremely efficiently computed.

New approached methods are under development for time-dependent problems.

3.3.- Comparison of Monte Carlo methods and FORM/SORM

The principal advantage of the Monte-Carlo method, but also of the majority of the variance reduction methods, is that they are valid for static, but also for dynamic models and for probabilistic models with continuous or discrete variables. Furthermore, there is no requirement on the failure functions – only the sign of the failure function is being used. In case of large computing times, it can be interesting to build an approximate mathematical model called response surface.

FORM and SORM are analytical and approximate methods, and their accuracy is generally good for small probabilities ($< 10^{-3}$). The analytical properties enable the methods to yield relatively inexpensive sensitivity factors. The basic random variables must be continuous, and the failure function must be continuous. With the optimisation procedures presently used in most cases, the failure functions should be smooth.

For small order probabilities FORM/SORM are extremely efficient compared to simulation methods, if the number of random variables is not too high. The CPU-time is for FORM approximately linear in the number of basic variables n, and the additional CPU-time for a SORM computation grows approximately with n^2. The absolute computation time depends on the time necessary to evaluate the failure function. This time may in fact depend on the actual values of the basic variables. The CPU-time is independent of the probability level, assuming a constant time for evaluation of the failure function. Table summarizes the advantages and drawbacks of Monte-Carlo simulation and FORM/ SORM methods.

Table 1 : Comparison of the characteristics of reliability methods

Simulations	**FORM/SORM**
Results	***Results***
Failure probability Error on the estimation Probability distribution of the response	Failure probability Most influential variables (on the probability) Efficiency (depends on the number of random variables)
Assumptions	***Assumptions***
No assumptions on the random variables (discrete, continuous, dependency, etc.) No assumptions on the limit state function	Continuous random variables Continuous limit state function (more suitable for optimization step)
Drawbacks	***Drawbacks***
Computation costs (depends on the probability level)	No error on the estimation Global minimum is required, but it is necessary to obtain all the minima of the optimization problem.

We think it is very difficult to introduce FORM/SORM method for "repetive" assessment in a Level 2 PSA, because the robustness of the scheme should be verified before. And feedback experience shows that efficient optimization methods are sensitive to be used. But they could be of big interest if they are applied to submodels or parts of L2 PSA [Saignes 05].

4.- Sensitivity analysis

Sensitivity Analysis (SA) techniques may be used pursuing different objectives, all of them related to getting knowledge about the behaviour of the system studied, in other words, related to getting information about the input-output relation. For examples, they could provide guidance as to where to improve the state of knowledge in order to reduce the output uncertainties most effectively, to steer research and development efforts, or better understand the modeling or to obtain a good confidence in the results (potentially large uncertainties). [Saltelli 00] is a good introduction to the different methods of SA and their use in industry.

They may be divided into numerical and graphical techniques, and may in many cases be used simultaneously, using the same data set. There are many different SA techniques that may be divided according to different criteria. Let us divide them into the following types:

- Regression based techniques
- Non parametric statistics used to identify relations between regions of input parameters and output variables
- Analysis of variance (ANOVA) based techniques
- Distribution sensitivity techniques.

All these methods are "complementary" and could be used according the characteristics of the physical models and the statistics of interest. All these methods assume that the random input variables are statistically independent.

4.1.- Regression based techniques

These techniques are based on using a sample (inputs plus outputs), obtained either by random sampling, or LHS or stratified sampling, though they could also be based on data obtained when applying design of experiments concepts, and try to identify if there is any linear, monotonic or polynomial, possibly considering also interactions, relation between inputs and outputs. Main tools are:

* Pearson correlation coefficient, which identifies linear relations between one input parameter and one output variable
* Spearman rank correlation coefficient, which identifies monotonic relations between one input parameter and one output variable
* Partial correlation coefficients (PCC) and standardised regression coefficients (SRC), which identify linear relations between inputs and outputs taking into account the effect of correlation between inputs
* Partial rank correlation coefficients (PRCC) and standardised rank regression coefficients (SRRC), which identify monotonic relations between inputs and outputs taking into account the effect of rank correlation between inputs.

In all these cases, the validity of the indices obtained relies on the quality of the linear (or monotonic) approach adopted. If such hypothesis is not right, other more suitable techniques should be used. In the case of the PCCs and related statistic, a measure of the goodness of the linear (or monotonic) hypothesis is the coefficient of determination (R^2). Coefficients of determination

close to one indicate good global validity of the hypothesis, which means that the sensitivity indices are reliable in that case. Coefficients of determination close to zero indicate lack of validity of the hypotheses.

4.2.- Non-parametric statistics

There are several statistics commonly used in data analysis that may be oriented to study specific relations between different regions of input parameters and output variable. In some cases, there is no clear trend in the behaviour of an output variable with respect to a given input parameter, but there could be some close relation between, for example, the 10% lowest observations of one of them and the 15% of largest observations of the other one. This kind of relation, under some circumstances could be not easy to be detected by regression-based techniques, while they are easily detected by non-parametric statistics. They are also closely related to some graphic tools to be discussed in next point. In what follows there is a list of those amongst the most useful:

1. Mann-Whitney test
2. Smirnov test
3. Kruskal-Wallis test

4.3.- ANOVA based techniques

Analysis of variance is a statistical method that allows decomposing the variability (variance) of a data set into different contributors. In normal experiments, the contributors to variability are the main parameters affecting the results of the experiment, their potential interactions and pure experimental error, or residual variability (experimental noise). In the case of computer experiments the last contributor doesn't exist, or is almost irrelevant (it is just due to round errors). ANOVA decomposition for computer results is based on Sobol's theorem of decomposition or expansion of any integrable function $f(\mathbf{x})$ in the n-dimensional unit hypercube (K^n). This expansion was named '*expansion of a function into summands of different dimension*' or '*High Dimensional Model Representation*' (HDMR) by its author. As a result of such decomposition, ways to estimate the contribution of main effects and interactions of different order may be studied. There are three well-known techniques to estimate those contributions to the global variability:

* Correlation ratios. This is the cheapest of these techniques in computational term, at least for models with many input parameters, but it allows to estimate only the contribution of main effects
* Sobol's sensitivity indices[9]. This is the most expensive technique, but allows to compute all the contributions
* Fourier Amplitude Sensitivity test (FAST)

The general idea of these methods is to decompose the total variance of the response, noted D, into 2^{p-1} terms (p is the number of random variables). For instance, if there are 3 random variables:

$$D = D_1 + D_2 + D_3 + D_{12} + D_{13} + D_{23} + D_{123} \tag{13}$$

The generalization to a model with p inputs is given by:

$$D = \sum_{i=1}^{p} D_i + \sum_{1\le i<j\le p} D_{i,j} + \ldots + D_{1,2,\ldots,p} \tag{14}$$

The first order terms D_i gather the effect of the uncertainty on the input parameter x_i (taken independently) on the response variance. The terms $D_{i,j}$, $i \neq j$ gather the effect of the interaction between the parameters x_i and x_j. And so on.
In dividing the equation (13) by D:

$$1 = \sum_{i=1}^{p} S_i + \sum_{1 \leq i < j \leq p} S_{i,j} + \ldots + S_{1,2,\ldots,p} \tag{15}$$

where the terms S are called the sensitivity indices. The S_i terms evaluate the contribution of the parameter x_i taken independently, the $S_{i,j}$ terms evaluate the contribution of the interaction between parameters x_i and x_j.

The problem of these methods, and specially Sobol method [Archer 97], is that a good estimation of these indices requires a big number of calculations. Thus it is often necessary first to calculate a surrogated model or a response surface validated in the domain of variation of the random variables.

4.4.- Comments about sensitivity analysis methods

Classical sensitivity indexes like PRC, PRCC SRC… could easily be implemented in software, or the database defined by the sets of inputs and computed outputs could be easily post-processed in statistical software. But underlying assumptions should always kept in mind when the results are analyzed.
Methods like FAST and Sobol define the points for which calculations are required; a coupling between L2 PSA software and statistical software is then necessary. For these two last methods, a lot of calculations could be required according a given precision, then the use of surrogate models could be necessary.

We does not think it is possible to apply such methods to a full L2 PSA. But they could be of big interest if they are applied to submodels or parts of L2 PSA.

Sensitivity analysis permits to distinguish the influence of aleatory and epistemic uncertainties. This difference is important for the decision-maker, because the way for mitigating their impact is different. For example case of the most influential uncertainties are epistemic and uncertainty of the output is “unacceptable”, an input uncertainty could be better known (and then its influence reduced) with new experimental results. In case of the most influential uncertainties are aleatory and uncertainty of the output is “unacceptable”, in the most cases it is necessary to modify the design of the structure or to define an additional barrier for example.

5.- Conclusions

Dealing with uncertainties in PSA level 2 requires using a set of statistical techniques needed to assess input uncertainty, to propagate uncertainties in an efficient way, to characterize appropriately output uncertainty and to get information from computer code runs though the intelligent use of sensitivity analysis techniques. This document has summarised the main techniques available in scientific and technical literature to perform such tasks.

In our analysis, we have trying to give information about the interest, the capacities and the suitability of these methods and techniques for the purpose of Level 2 PSA and physic codes used in support. Our position about their implementation in L2 PSA software has been written; it could

be noticed that a lot of these methods are computing time consuming, and seems more suitable for the analyse of submodels or for focused about a question.

6.- References

[Archer 97] A. Archer & al. Sensitivity measures, anova-like techniques and the use of bootstrap. *Journal of statistical Computation and Simulation*, vol. 58, pp 99,120, 1997

[Baldeweck 97] H. Baldeweck. *Méthodes des éléments finis stochastiques. Applications à la géostatistique et à la mécanique de la rupture*. PhD University Evry-Val d'Essonne, 1997

[Box 87] G.E.P Box and N.R Draper. *Empirical Model Building and Response Surface,* J. Wiley & Sons, Wiley Series in Probability and Mathematical statistics, 1987

[Chevalier 97] M. Chevalier, D. Perez, J-M. Biasse and J-M. Mitterand. Resampling and simulation on small-size samples. in C. Guedes Soares (Ed), *Advances in Safety & Reliability*, Pergamon, 1997

[Devictor 99] N. Devictor, S. Chapuliot and M. Eid, Monte-Carlo simulation to assess structures reliability. in *Proceedings of Esrel'99*, Edinburgh, 1999

[Devictor 00] N. Devictor and J-M. Martinez. Nonlinear regression methods in uncertainty and sensitivity studies and reliability computations. in *Proceedings of ESREL 2000*. 2000

[Devictor 04] N. Devictor. Advances in methods for uncertainty and sensitivity analysis. *Proceedings of the workshop "Level 2 PSA and Severe Accident Management", OCDE/AEN/CSNI/WGRISK*, Köln, March 2004

[Hastie 02] T. Hastie, R. Tibshirani and J. Friedman. The elements of statistical learning. Springer, 2002

[Helton 04] J. Helton and W. Oberkampf (Editors). Alternative representations of epistemic uncertainty. Reliability Engineering and System Safety, 84, 2004.

[Iooss 05] B. Iooss, F. Van Dorpe and N. Devictor, Response surfaces and sensitivity analyses for an environmental model of dose calculations. *Reliability Engineering and System Safety*, in press, 2005

[Lannoy 01] A. Lannoy and H. Procaccia. *L'utilisation du jugement d'expert en sûreté de fonctionnement*. Edition Lavoisier Tec&Doc, Collection Recherche et Développement d'EDF, 2001

[Melchers 99] R.E. Melchers. *Structural reliability analysis and prediction*. J.Wiley & Sons, 1999

[Mengolini 05] A. Mengolini, K. Simola and R. Bolado. *Formal Expert Judgment. An overview.* EUR 21772 EN. 2005.

[NEA 94] NEA/CSNI Report of a workshop on uncertainty analysis methods. NEA/CSNI/R(94)20 (Vol. 1 and 2), 1994

[NEA 97] NEA/CSNI Report on the uncertainty methods study for advanced best estimate thermal hydraulic codes applications. NEA/CSNI/R(97)35 (Vol. 1 and 2), 1997.

[NEA 99] NEA/CSNI Report on best estimate methods in thermal hydraulic safety analysis. NEA/CSNI/R(99)22, 1999.

[Pendola 00] M. Pendola. *Fiabilité des structures en contexte d'incertitudes statistiques et d'écarts de modélisation*. PhD. Blaise Pascal University, Clermont-Ferrand, 2000

[Pearson 65] E.S. Pearson and M. Tukey. Distributions whose first four moments are known. *Biometrika*, 6, pp 133-137, 1965

[Sacks 89] J. Sacks, W.J. Welch, T.J. Mitchell and H.P. Wynn, Design and analysis of computer experiments. *Statistical Science*, 4:409-435, 1989

[Saltelli 00] A. Saltelli A., K. Chan et E.M. Scott E.M. (Eds). *Sensitivity Analysis*. J. Wiley, Wiley Series in Probability and Statistics, 2000

[Saporta 90] G. Saporta. *Probabilités analyses des données et Statistiques*. Editions Technip, Paris, 1990

[Wallin 84] K. Wallin, The Scatter in K_c-Results. *J. Engineering Fracture and Mechanics,* 19, 1984

[WEJ 05] Proceedings of the "Workshop on the use of expert judgment in decision making process". Aix-en-Provence, June 2005, in press.

[Wilks 41] S. Wilks. Determination of sample sizes for setting tolerance limits; *Ann. Math. Statist*., 12, pp. 91-96, 1941