

**WORKSHOP ON THE EVALUATION OF UNCERTAINTIES IN RELATION TO  
SEVERE ACCIDENTS AND LEVEL II PROBABILISTIC SAFETY ANALYSIS  
CADARACHE, FRANCE, 7-9 NOVEMBER 2005**

**TITLE**

The use of Monte-Carlo simulation and order statistics for uncertainty analysis  
of a LBLOCA transient (LOFT-L2-5)

Authors :

Eric CHOJNACKI IRSN/DPAM/SEMIC

Jean-Pierre BENOIT IRSN/DSR/ST3C

IRSN : Institut de Radioprotection et de Sûreté Nucléaire

**ABSTRACT**

Best estimate computer codes are increasingly used in nuclear industry for the accident management procedures and have been planned to be used for the licensing procedures. Contrary to conservative codes which are supposed to give penalizing results, best estimate codes attempt to calculate accidental transients in a realistic way. It becomes therefore of prime importance, in particular for technical organization as IRSN in charge of safety assessment, to know the uncertainty on the results of such codes.

Thus, CSNI has sponsored few years ago (published in 1998) the Uncertainty Methods Study (UMS) program on uncertainty methodologies used for a SBLOCA transient (LSTF-CL-18) and is now supporting the BEMUSE program for a LBLOCA transient (LOFT-L2-5). The large majority of BEMUSE participants (9 out of 10) use uncertainty methodologies based on a probabilistic modelling and all of them use Monte-Carlo simulations to propagate the uncertainties through their computer codes. Also, all of 'probabilistic participants' intend to use order statistics to determine the sampling size of the Monte-Carlo simulation and to derive the uncertainty ranges associated to their computer calculations.

The first aim of this paper is to remind the advantages and also the assumptions of the probabilistic modelling and more specifically of order statistics (as Wilks' formula) in uncertainty methodologies. Indeed Monte-Carlo methods provide flexible and extremely powerful techniques for solving many of

the uncertainty propagation problems encountered in nuclear safety analysis. However it is important to keep in mind that probabilistic methods are data-intensive. That means, probabilistic methods cannot produce robust results unless a considerable body of information has been collected. A main interest of the use of order statistics results is to allow to take into account an unlimited number of uncertain parameters and, from a restricted number of code calculations to provide statistical tolerance limits for any code results. A proof and an extension of this statistical theorem will be given. From this proof, it will appear easily why the use of order statistics results requires the Simple Random Sampling method (SRS).

The second aim of this paper is to illustrate the benefit of these techniques from the application of the IRSN uncertainty methodology on the transient LOFT-L2-5. To achieve this aim, we will use the results obtained in the frame of our participation to the BEMUSE program to clarify how to perform and analyse a Monte-Carlo simulation. In particular, it will be shown how order statistics provide valuable results for estimating percentiles of relevant safety quantities.

Finally, from our experience gained during BEMUSE project, we will conclude on the applicability of Monte-Carlo simulation to derive uncertainty ranges for safety purposes.

## **Introduction**

To demonstrate that the nuclear power plants are designed to respond safely at numerous postulated accidents computer codes are used. The models of these computer codes are an approximation of the real physical behaviour occurring during an accident. Moreover the data used to run these codes are also known with a limited accuracy. Therefore the code predictions are not exact but uncertain. To deal with these uncertainties, safety demonstration can follow two different ways. The first way is to use conservative codes with conservative data. Such studies contain deliberate pessimisms and unphysical assumptions. It is then argued that the overall predictions are worse than reality. The second way is to use 'best estimate' codes with 'best estimate' input data to obtain a best estimate calculation. If such calculations are performed for safety studies, it is necessary to value (or to overvalue) the uncertainty associated to this estimation.

The use of best-estimate codes is motivated by both safety and economical reasons. First, the conservatism of results issued from conservative codes is not

simple to prove due to presence of counter-reactions. Moreover, the use of best-estimate codes allows to improve accident management procedures thanks to a better understanding of accident progress. Secondly, for economical reason, it is expected that the use of best-estimate codes will allow to relax technical specifications and core operating limits set by conservative calculations.

Thus, uncertainty methods are needed to use best-estimate codes for safety purposes and to evaluate safety margins around the best-estimate values. Among uncertainty methods, two main kinds of methods can be distinguished: deterministic or probabilistic methods.

In deterministic methods, the uncertainty intervals around the best-estimate values depend on the analyst. Indeed, in deterministic methods, the investigation of the uncertainty domain, i.e. the choice of combinations of values, requires the analyst's decision. Moreover, in deterministic methods it is very difficult to quantify the confidence associated to the produced uncertainty results. On contrary, the probabilistic methods allow to take into account in a natural way the likelihood associated to the combined values of the identified uncertainty sources. The measure of the confidence associated to the results is also statistically established.

In the first section of this paper, we will recall the principle and advantages of the probabilistic modelling. The second section will present in detail the use of order statistics in Monte-Carlo simulations. Then, we will describe the LOFT L25 experiment and the main results of our uncertainty analysis using order statistics.

## **1. Probabilistic Modelling : Principle and Advantages**

The probabilistic approach as any uncertainty propagation method, requires to identify all the potentially important contributors to uncertainty of code results. These contributors are generally referred as the 'uncertainty sources' or as the 'uncertain parameters'.

Then, it is necessary for each of uncertain parameters to quantify its own uncertainty. In probabilistic methods, that step requires to select a probability distribution which represents the quantification of the confidence associated to the respective values. If dependencies are known between uncertain parameters and judged to be potentially important, then these dependencies need to be quantified. In this way, each value of the uncertainty domain is weighted by its likelihood. The probabilistic methods consist to evaluate from this knowledge the likelihood associated to each possible results.

The propagation of the uncertainties from the uncertainty sources to the results through the computer code, requires, excepted for very simple codes, a numerical estimation. This estimation is obtained thanks to a Monte-Carlo simulation. In Monte-Carlo simulation, a model is run repeatedly, using different values for each of the uncertain parameters each time. The values of each of the uncertain parameters is based on its probability distribution. In this way, one value for each uncertain parameter is sampled simultaneously in each repetition of the simulation. The results of a Monte-Carlo simulation (100, 1000 or more sets of samples of the uncertainty parameters) lead to a sample of the same size for each code response. This sample can be used to get any typical statistics such as mean or variance and to determine the cumulative distribution function (CDF). The CDF allows to get quantitative insights regarding the percentiles of the distribution. Therefore the estimation of the CDF is quite important for safety issues.

It can be noticed<sup>[1]</sup> that the calculation of any of these statistics can be computed in practice as an average of an arbitrary function  $g$ . Indeed, if we consider the sample mean of  $g(Y)$  :  $\frac{1}{N} \sum_{i=1}^N g(Y_i)$ , we can remark by taking respectively for example  $g(Y)=Y$ ,  $g(Y)=(Y - \bar{Y})^2$  or  $g(Y) = 1_{]-\infty, x]}(Y)$  that the sample mean of  $g(Y)$  is respectively the sample mean of  $Y$ , the variance of  $Y$  and the cumulative distribution of  $Y$  in  $x$ . The Monte-Carlo simulation is therefore a quite simple way to obtain useful statistics about the distributions of code outputs.

The advantage of Monte-Carlo methods with respect to the deterministic uncertainty analysis methods is that the combination of uncertain parameters values are performed in such a manner that they allow to quantify the likelihood of any particular code result.

The aim of the different Monte-Carlo sampling techniques is to speed up the convergence of the average of  $g(Y)$  toward the desired statistical estimator. Among Monte-Carlo methods, in thermal hydraulics, two random sampling methods are generally chosen : the Simple Random Sampling method (SRS) and the Latin Hypercube Sampling method (LHS). The SRS method is the most obvious random sampling method. For each uncertain variable, a sample is generated independently according to its probability distribution. The LHS method requires to draw values for each uncertain parameter in all portions of its probability distribution. The principle of LHS sampling method<sup>[2]</sup> is to divide each probability distribution into  $N$  strata of equal marginal probability  $1/N$ , and to sample once for each stratum. A practical way to model the potential

dependencies between parameters is due to Iman & Conover. Iman & Conover described in their paper a procedure which allows to induce the wished rank correlations between parameters and this procedure is applicable whatever were the marginal distributions may be. The Iman & Conover procedure is applicable both for SRS and LHS methods and is widely used in Monte-Carlo simulations. One advantage of the LHS method appears when the code output is dominated by only a few of the uncertain parameters. Indeed, this method ensures that each of the uncertain parameters is represented in a fully stratified manner, no matter which components might turn out to be important. On contrary, if the relationship between uncertain parameters and code outputs is more complex (non monotonicity with many parameters) the SRS method can lead to better estimation of the probability distribution associated to the code output. Moreover, as the LHS method is not a pure random method, the direct estimations of the percentiles provided by order statistics theorems cannot be applied.

## 2. Use of Order Statistics<sup>[3][4]</sup>

The principle of order statistics is to derive statistical results from the ranked values of a sample. In particular order statistics allow to get direct estimations of the percentiles and also their confidence intervals.

### 2. 1 Use of order statistics from a single draw

A well known probabilistic result is that the random variable  $Y$  defined from the random variable  $X$  as the value of the CDF of  $X$ , noted  $F_x$ , applied to  $X$ . i.e. :  $Y=F_x(X)$  follows an uniform distribution.

This result means that the probability for any particular occurrence of a random variable to be lower (resp. upper) than it's  $\alpha$ -percentile is  $\alpha$  (resp.  $1-\alpha$ ).

#### Proof :

Let be  $X_\alpha$  the  $\alpha$  percentile of the random variable  $X$ .  $X_\alpha$  is a deterministic value (and eventually unknown value if  $F_x$  the CDF of  $X$  is not known).

$$\begin{aligned} \forall \alpha \in [ 0. , 1. ], \quad \text{Proba}(Y \leq \alpha) &= \text{Proba}(F_x(X) \leq \alpha) \\ \Leftrightarrow \text{Proba}(X \leq X_\alpha) &= \text{Proba}(X \leq F_x^{-1}(\alpha)) = F_x(F_x^{-1}(\alpha)) = \alpha \end{aligned}$$

#### Example :

If we have an unique occurrence of a random variable, then it can be concluded that the probability that this observed value to be upper than the percentile 95% is lower than 5%.

## 2.2 Use of order statistics from the $k^{\text{th}}$ out of $N$ draws

The former result can be easily generalized when we have a sample of values instead of a single occurrence of a random variable. Let us consider a sample of size  $N$  of a random variable  $X$  and call  $x_{(1)} \leq \dots \leq x_{(N)}$  the sorted values of this sample. The observed values  $x_{(1)} \leq \dots \leq x_{(N)}$  can be viewed as the particular occurrence of a random sorted vector  $(X_{(1)}, \dots, X_{(N)})$ . Now let us prove that the random variable  $Y$  defined as the value of the CDF of  $X$ , noted  $F_X$ , applied to  $X_{(K)}$ . i.e.  $Y = F_X(X_{(K)})$  follows a beta law of parameters  $K$  and  $N-K+1$ .

Proof :

Let be  $X_\alpha$  the  $\alpha$  percentile of the random variable  $X$ . As we have already mentioned, we have by definition of  $X_\alpha$ :

$$X_\alpha = F_X^{-1}(\alpha) = \forall \alpha \in [0, 1.]$$

Now consider the random variable  $X_{(K)}$ . The probability that  $X_{(K)}$  is less than  $x$  is equal to the probability to have at least  $K$  values of the sample less than  $x$ . For each draw its probability to be lower (resp. upper) than  $x$  is obviously (by definition of the CDF)  $F_X(x)$  (resp.  $1 - F_X(x)$ ) Thus, we can write, as far as all the draws are randomly and independently generated :

$$\text{Proba}(X_{(K)} \leq x) = \sum_{i=k}^n C_n^i F^i(x) (1 - F(x))^{n-i}$$

We can now remark that if  $x = X_\alpha$  the right part of the above equality does not depend on  $F_X$ . Indeed, we obtain :

$$\text{Proba}(X_{(K)} \leq X_\alpha) = \sum_{i=k}^n C_n^i \alpha^i (1 - \alpha)^{n-i}$$

Now consider a random variable  $Y$  following a beta law of parameters  $k$  and  $n-k+1$  By definition of the random variable  $Y$ , its density distribution is :

$$f_y(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1} (1-x)^{n-k} \text{ if } 0 \leq x \leq 1 \text{ or } 0 \text{ otherwise.}$$

Consequently the CDF of  $Y$  in  $x$  (for  $0 \leq x \leq 1$ ) is given by  $F_Y(x)$  :

$$F_Y(x) = \int_0^x \frac{n!}{(k-1)!(n-k)!} t^{k-1} (1-t)^{n-k} dt$$

If we integrate by parts  $F_Y(x)$ , we have :

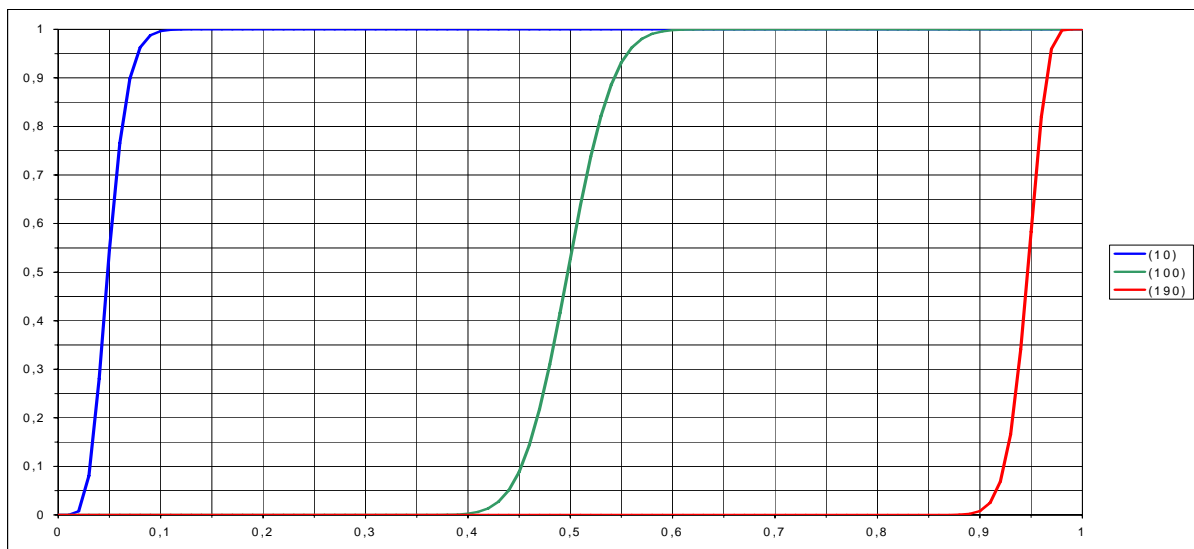
$$F_Y(x) = \int_0^x \frac{n!}{(k-1)!(n-k)!} t^{k-1} (1-t)^{n-k} dt = \sum_{i=k}^n C_n^i x^i (1-x)^{n-i}$$

This last equality proves that  $F_X(X_\alpha) = F_Y(\alpha)$  ; i.e. the random variable  $F_X(X_{(K)})$  follows a beta law of parameters  $K$  and  $N-K+1$ .

Consequently, the value of the cumulative distribution of  $X_{(k)}$  in  $X_\alpha$  does not depend of the distribution of  $X$ , so that it is now possible to derive confidence intervals for any percentiles directly from the sample values without having to determine the probability distribution of the random variable.

**Example 1 :** How to use the 10<sup>th</sup>, 100<sup>th</sup> and 190<sup>th</sup> draw out of 200 draws to estimate the lower, likely or upper value of different percentiles ?

The former statistics order theorem allows us to know for a random variable  $X$  the probability for each sorted value, denoted  $(i)$ , out of a random sample of size  $n$  to be lower than its percentile  $X_\alpha$  without to estimate the CDF of  $X$ .



On the above figure are plotted the CDFs of beta :  $\beta_{(10,191)}$ ,  $\beta_{(100,101)}$  and  $\beta_{(190,11)}$ . For example, the 'blue' curve (i.e. the CDF of  $\beta_{(10,191)}$ ) shows that the 10<sup>th</sup> draw has a probability lower than 0.7% to be lower than the percentile 2% of a random variable  $X$  for which we observed a sample of size 200. Similarly we derive from the blue curve, that the 10<sup>th</sup> draw, has a probability about 50% to be lower than the percentile 5% and has a probability more than 99.9% to be lower than the percentile 11%. These results did not require to know the probability density of  $X$ . The green and red curves allow to derive similar results from the 100<sup>th</sup> and 190<sup>th</sup> draw of this sample.

**Example 2 :** How to use order statistics to determine the minimum sample size which allows to derive an upper limit of the percentile  $\alpha$  at a probability  $\beta$  ( $\beta$  is also named the confidence level) ?

To determine this minimum sample size, it is enough to notice that this minimum sample size  $n$  is obtained when we take the largest value of the sample (i.e:  $X_{(n)}$ ) as the upper limit of the percentile.

$$k=n \Rightarrow \beta_{(n,1)}(x) = n x^{n-1} \Rightarrow \alpha^n \leq 1-\beta \Leftrightarrow n \geq \ln(1-\beta)/\ln(\alpha)$$

The following table gives numerical values of  $n$  for different percentiles at different confidence levels.

*Table : minimum sample size*

$\alpha \backslash \beta$	0.80	0.90	0.95	0.99
0.80	8	16	32	161
0.90	11	22	45	230
0.95	14	29	59	299
0.99	21	44	90	459

**Example 3** : How to evaluate the sample size effect on the accuracy of estimated percentiles ?

Let us consider again the example 1. We suppose that we already get a random sample of size 200, and we are interested by the estimation of the percentile 95%, noted  $X_{95\%}$ . Obviously, the best-estimate of this percentile is the value observed for  $X_{(190)}$ . From the previous theorem, it is possible to know that :

$$\text{Probability}( X_{(184)} > X_{95\%} ) = 2.4\% \text{ and Probability}( X_{(196)} < X_{95\%} ) = 2.6\% .$$

Therefore, it is very likely (95%) that the unknown value of the percentile 95% is between the two observed values  $X_{(184)}$  and  $X_{(196)}$ . An increasing of the sample size will allow to reduce our inaccuracy on the percentile 95%, i.e. the difference between the two observed values  $X_{(184)}$  and  $X_{(196)}$ . Thus, for example if we consider a sample of size 400, we have :

$$\text{Probability}( X_{(368)} > X_{95\%} ) = 0.4\% \text{ and Probability}( X_{(392)} < X_{95\%} ) = 0.2\% .$$

### 3. Description of LOFT L2-5

The LOFT facility simulated the major components and the system responses of a commercial PWR during a loss-of-coolant accident (LOCA). The core was a semi-scale one with an active height of 66 inches. The experimental assembly

included five major subsystems which were instrumented such that system variables can be measured and recorded.

The L2-5 experiment has been successfully completed on June 16, 1982 in the Loss-of-Fluid Test (LOFT) facility at INEL (Idaho National Engineering Laboratory). This experiment simulated a guillotine rupture of an inlet pipe in a pressurized water reactor with a true nuclear core. The experiment L2-5 was initiated, after operating the reactor at 36.0 MW for 40 effective full power hours to build up a fission decay product inventory, by opening two quick-opening blowdown valves upstream a blowdown suppression tank simulating the reactor containment behaviour.

The following objectives were defined for experiment L2-5 :

1. determine if early core rewet occurs following a 200% double-ended cold leg break with immediate primary coolant pump trip,
2. provide data on core thermal response which can be used to evaluate computer code predictions and to compare with acceptance criterion,
3. determine system behaviour and core thermal response during the reflood portion of a double-ended cold leg break experiment,
4. evaluate cladding surface thermocouple effects during blowdown and reflood by comparing the responses of LOFT fuel bundle instrumentation.

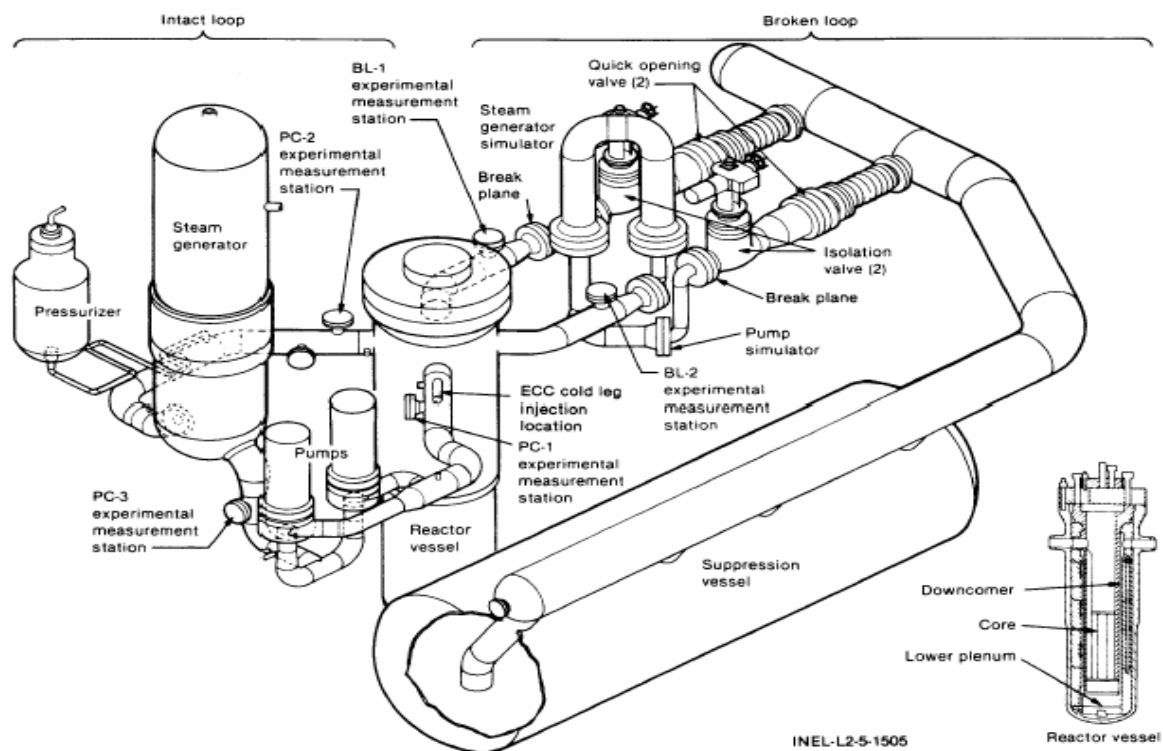


fig 1: View of the experimental LOFT system

Since the L2-5 experiment, several computations, such as ISP13, have been conducted to simulate the behaviour of the LOFT system and compare with the experimental results. In the same way, as first phase (2003-2005) of the OECD/CSNI program on Best Estimate Methods for Uncertainty and SEnsitivity analysis (BEMUSE), the L2-5 experiment has been chosen to apply uncertainty methodologies on a Large Break Loss Of Coolant Accident (LB-LOCA transient) performed on an integral test facility. The second phase of BEMUSE program (2006-2007) will consist to apply these methodologies to a LB-LOCA transient postulated on a nuclear power plant. We show hereafter some results obtained at IRSN in the first program phase.

#### 4. Results

To illustrate the use of order statistics and the ability of the method to provide statistical tolerance limits for any code results, we present the results of 383 runs performed with 27 uncertain parameters sampled following a SRS technique.

The results shown in fig. 2 represent the peak cladding temperatures (PCT) of a hot rod in a hot channel. It is important to remind that the computed temperatures, except for the reference calculation, are ranked (or ordered) at each time, so that the 'ranked' curves represent for each time step the 355<sup>th</sup>, 364<sup>th</sup> and 372<sup>nd</sup> largest values and do not represent actual evolutions of the PCT.

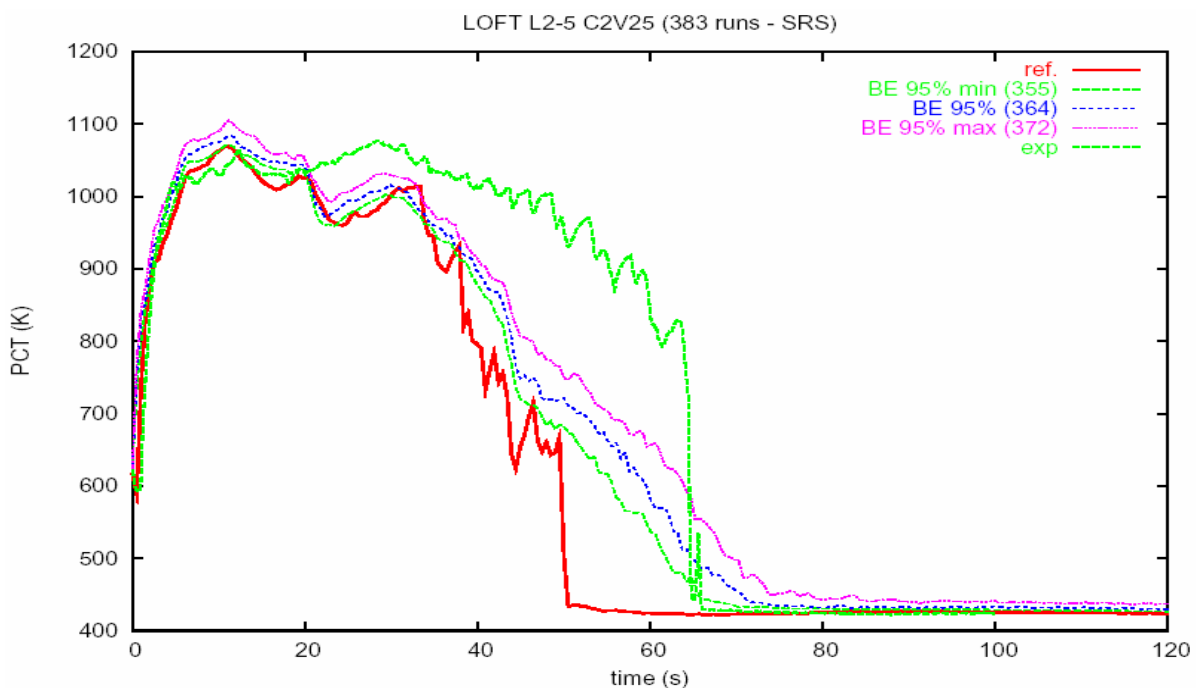


Fig 2: Peak Cladding Temperature of a hot rod in a hot channel: 383 runs

Hereafter, table 1 show the results from 496 runs performed with 27 uncertain parameters sampled following a SRS technique.

Tab 1: Peak Cladding Temperature of a hot rod in a hot channel: 496 runs

PCT (°K)	5% percentile	95% percentile
Min	926	1076
BE value	935	1087
Max	939	1105
Reference value	1069	
Experimental value	1077	

The 'BE 95%' curve represents the best estimation of the percentile 95% which can be derived from a random sample of 383 results. The gap between the 'BE 95%min' and 'BE 95% max' curves represents the imprecision (at the 95% confidence level) of the estimation given by the 'BE 95%' curve due to the limited sample size. For example, it can be read directly from the figure 2 or more easily from the table 1 that the best estimation value of the percentile 95% of the peak clad temperature is 1087K. Taking into account the limited sample size, the likely lower value of the percentile 95% is 1076K and its likely upper value is 1105K. Therefore, the limited sample size leads very likely (at 95%) to an uncertainty of ~30K ( $=1105K-1076K$ ) on the 95-percentile of the peak clad temperature ( $=1087K$ ). This uncertainty analysis shows also that the discrepancy observed after 20s of transient between the experiment and CATHARE calculations cannot be explained by the choice of the uncertain parameters and their associated PDFs taken to model our state of knowledge . One can also observe that the reference calculation leads to values near to the 95-percentile. This is due to the choice of reference values taken in the reference calculations with respect to the choice of PDFs taken for uncertain parameters.

## Conclusion

The use of order statistics in Monte-Carlo simulations provides both an extremely simple and powerful way to evaluate any typical statistics as the 95-percentile and their associated confidence intervals. Moreover the width of the confidence intervals allows to quantify the effect due to the limited size of the observed sample. The uncertainty margins derived from order statistics do not require any assumptions neither on the number of uncertain parameters nor on

the relationships between the code results and the uncertain parameters. Consequently, the provided results are independent of all these assumptions. However, it is quite important to remind that as for any Monte-Carlo simulation, the quality of results fundamentally depends on the selected uncertain parameters list and on the relevance of the choice of their PDFs to model our state of knowledge.

References:

[1]M.D. Mc Kay, R.J. Beckman, W.J. Conover  
A Comparison of Three Methods for Selecting Values of Input Variables in the  
Analysis of Output from a Computer Code  
Technometrics , Vol. 21, N°2, May 1979

[2]S.P. Millard, N.K. Neerchal  
Environmental Statistics with S+  
Applied environmental Statistics , CRC PRESS 2001

[3]J.P. Lecoutre, P. Tassi  
Statistique non paramétrique et robustesse  
Economica 1987

[4]W.J. Conover  
Practical Nonparametric Statistics  
Wiley Series in Probability and Statistics 1999