Sensitivity strategy supporting the estimate of extremely low probabilities

Cédric J. Sallaberry[a*], Robert E. Kurth[a]
[a] Engineering Mechanics Corporation of Columbus (Emc^2)
Columbus, OH, USA

Abstract: The probabilistic analysis of possible piping rupture or significant loss of coolant accident (LOCA) is a complex problem as it involves many mechanisms and generates low to extremely low probabilities of events. This topic is of particular interest in the nuclear industry and a conjoint effort between the US NRC and EPRI over the last 10 years has led to the development of the Extremely Low Probability of Rupture (xLPR) code to assess probability of rupture in nuclear piping systems. In this paper we focus on using the code in conjunction with statistical approaches to increase confidence in the results, which is a necessary aspect of any risk-informed approach. One of the step is to understand the uncertainty in the output of interest and identify the most influential parameters responsible for this uncertainty. We present here a strategy based on a suite of regression techniques to rank the parameters importance that can be used in conjunction to sampling and importance methods (in an iterative fashion) in order to converge to stable and well understood results.

Keywords: Sensitivity Analysis, Regression Analysis, Rare events, xLPR

1 INTRODUCTION

Despite the aging of nuclear infrastructures and resulting uncertainties, they are required to perform in a highly reliable manner. Of particular interest is the assessment of piping rupture or a possible significant loss of coolant accident (LOCA) in a nuclear power plant. Particularly, the occurrences and evolution of cracks that could occur due to several conditions (Primary Water Stress Corrosion Cracking or Fatigue) on welds that join two dissimilar or similar metal pipes is considered as a potential risk leading to such adverse condition. While the US NRC does not regulate on risk it does relies on risk-informed approach in order to understand the consequence of uncertainty and support the conclusion drawn.

To this end the US NRC, in conjunction with EPRI, has developed the Extremely Low Probability of Rupture (xLPR) code to assess probability of rupture in nuclear piping systems [1] [2]. This code models the likelihood and evolution of potential cracks in the weld. It considers several mechanisms and plant properties including crack initiation, growth, coalescence and stability, weld residual stresses and materials properties. The code also considers potential human intervention such as mechanical and/or chemical mitigation as well as the impact of in service inspections and leak detection. Due to the defense in depth approach, and the regular inspection schedule implemented coupled with constant measurement to detect potential leakage, the probability of rupture or other adverse events such as LOCA is expected to be extremely low (in the order of 10^{-5} or lower) that may render direct Monte Carlo method impractical.

In a proposed joint paper, we will discuss different sampling and optimization strategies that are required to estimate, with confidence, the probabilities of failure events. When these probabilities are extremely low, one has often to rely on optimization methods in order to concentrate on the area of the input space that cover those adverse events. Such approach requires one to know a priori which input variables are important (in terms of uncertainty) and how they affect the selected output of interest.

* *Csallaberry@emc-sq.com*

One of the purposes of sensitivity analysis is to determine such influences and rank the uncertain parameters [3]. As such, this paper presents strategy based on a range of regression methods, to estimate the impact of input uncertainty, including non-monotonic and conjoint influence [4]. The latter is fairly common in complex systems such as the ones considered by the xLPR code, involving many mechanisms and the interaction of influential random variables. Furthermore, while the physics studied is usually monotonic, the addition of mitigation techniques and inspections introduce non-monotonicity (the most obvious failures will be detected and repaired). The results of the regressions are then combined with expert elicitation to better understand the impact of uncertainty and focus on the most important contributors: when a large number of potentially uncertain inputs is considered (up to 300 or more) it is often necessary to revisit the drivers of the problem to increase confidence in the results.

Furthermore, once the important input parameters have been identified, a handful can be selected as candidates for importance sampling or adaptive sampling in order to estimate these extremely rare events.

2 CONTEXT OF THE ANALYSIS

The problem we are investigating numerical simulation of complex physics in engineered systems composed of many inputs and interwoven mechanisms. Furthermore, while the consequences of the events under consideration are serious, they represent extremely rare events.

The physics is usually fairly monotonic (i.e, the events happens only if a high or low threshold is met) however with strong conjoint influence expected. Furthermore human interaction, either as a response (such as inspection of the system, detection of potential issues via sensors) or preemptively (such as mitigation), introduces potential non-monotonicity and disjoint response spaces in the areas of interest (as most obvious issues are usually quickly detected)

Some of the inputs may affect different parts of the analysis, in a different way. As a reults, extreme values can be both beneficial and detrimental depending on the sub-model considered which makes it difficult to assess prior yo performing the analysis the overall impact of such extreme values.

Due to the large uncertainties present in the system whether in the inputs considered or the model used, a probabilistic approach is needed to support any risk informed decision making. However these complex models are usually expensive in term of running time and memory requirements, which limits the number of runs that can be performed in a Monte Carlo approach.

All of these requirements and constraints underline the need for an efficient method to identify the most important factors in order to focus on what is really driving the uncertainty of the output of interest and be able to optimize the analysis toward these factors.

3 REQUIREMENTS FOR THE SENSITIVITY ANALYSIS

Sensitivity Analysis (SA) is the analysis of the importance of the input uncertainty on the output uncertainty [5] [6]. It is traditionally used to rank the uncertain inputs and parameters by importance, either to reduce this uncertainty or to select where importance sampling should be applied to increase confidence in the results.

Following the constraints listed in the previous section, the SA technique needs to have certain features in order to be efficient:

- It needs to be able to capture different patterns, both monotonic and non-monotonic

- It also needs to capture the influence if input interactions which we call the conjoint influence
- It should be able to handle large number of uncertain inputs (as high several hundreds)
- Due to the rare nature of the events, the method has to be efficient and robust even when few realizations lead to the adverse condition
- It also as to be fast enough so it can be applied to a large range of outputs (surrogate/intermediate outputs may have to be considered) if no adverse events occur.
- It has to be strongly adaptable. In our experience there is no universal approach that would work for any case, thus the approach needs to be flexible enough so it can change when needed.

4 METHODOLOGY

The method we propose relies on a suite of regression techniques [4]. They are run for each output of interest to estimate the influence of the input uncertainty toward this output. Results are then aggregated into a single number used to rank the inputs by their importance. The quantitative result are supported by scatterplots which are more qualitative and visual.

4.1. Selected regression techniques

4.1.1 Linear or Rank Regression

The linear regression is one of the oldest most known and widely used regression technique. The rank regression is a linear regression on rank transformed input and output variables under consideration. The form of the final regression model is additive (no conjoint influence) and as follow:

$$Y = a_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n = a_0 + \sum_{i=1}^{n} a_iX_i + \varepsilon = \bar{Y} + \varepsilon \quad (1)$$

where ε represents (for this regression and the subsequent ones) the difference between the output of interest Y its estimate $\bar{Y}$.

When the number of inputs is large, stepwise regression is a useful technique as it adds the input variables one at a time with an inclusion and exclusion criteria. It thus avoids overfitting the model by including more input variables than necessary. The stepwise approach starts with trying to find the best fit with only one parameter by testing all possible input parameters. It then builds up from this initial fit by selecting the best fit with two parameters, conditional upon keeping the first parameter, and so on. A stopping criterion is set via a generalized cross validation approach. Rank regression is effective in capturing monotonic relationships between inputs and outputs. The non-parametric aspect makes it less sensitive to outliers (which may be desired or not desired depending on this example). This technique is limited to additive models where no conjoint influences are considered and may perform poorly on non-monotonic relationships.

Three metrics are used for each input variable included in the rank regression results. Two are based on the coefficient of determination, noted conventionally R^2, which represents the amount of variance explained by the regression model. The coefficient of determination is a normalized value which varies between 0 (no variance explained) and 1 (all the variance explained).

- $R^2_{inc}(i)$ reports the cumulative coefficient of determination of the rank regression model when the i^{th} variable has been added (that includes all variables up to the i^{th} for the model).

- $R^2_{cont}(i)$ reports the incremental gain in R^2 when the i^{th} variable has been added compared to the model with $(i-1)$variables. While this value does not correspond exactly to the fraction of variance in the output explained by the variance in the i^{th}input, it remains a good estimate.
- $SRC(i)$ reports the value of the standardized rank regression coefficient for i^{th} variable. This value, also an indicator on the strength of the linear relation (or monotonic relation – we talk then of SRRC for standardized *rank* regression coefficient) between the i^{th} variable and the output of consideration. This also indicates the direction of the linear trend either positive (high values of input associated with high values of output) or negative (high values of input associated with low values of output).

4.1.2 Response surfaces

The other regressions are treated differently, as their models do not allow for a direct estimate of the contribution to each individual input to the variance of the output. As a result, they are used as response surfaces first with their quality estimated based on the coefficient of determination (R^2) of the final model. This model is then used to generate a large number of realizations to estimate the variance decomposition using the Sobol decomposition. This technique estimates the contribution of each input and their potential interactions (i.e., conjoint influence) via an integral decomposition of variance [17]. It requires a large number of realizations (tens to hundreds of thousands) to be accurate and cannot usually be applied directly to the computationally demanding models that we consider. The regression techniques however create analytical models that can be run a large number of times relatively quickly (seconds or minutes). The Sobol decomposition assesses the importance of each variable (according to its uncertainty) to the variance of the output considered. The quality of the estimate is strongly dependent on the quality of the regression model and caution should be applied when the R^2 value is relatively low.

The Sobol decomposition leads to different measures of effect than those used in stepwise linear regression. The two metrics selected for this analysis are described below:

- S_i, the first order sensitivity index for variable i, characterizes how much of the variance of the selected output is explained by the input parameter under consideration by itself (i.e., without conjoint influence). This index therefore can be assimilated to the R^2_{cont} from the linear or rank regression approach.
- T_i, the total order sensitivity index for variable i, characterizes how much of the variance of the selected output is explained by the input parameter alone *and its interaction* with the all the other uncertain inputs (i.e., conjoint influence). It has no analogue in the rank regression model as the additive model does not capture conjoint influences.
- The difference $(T_i - S_i)$ provides an estimate of the conjoint influence for a single input on the output considered.

4.1.3 Recursive partitioning

Recursive partitioning regression is a regression method based on regression tree techniques. A regression tree splits the data into subgroups in which the values are relatively homogeneous. The regression function is constructed using the sample mean of each subgroup. This approach results in a piecewise constant function over the input space under consideration. The predictive model is of the form:

$$Y = \sum_{s=1}^{nP} \left(d_s I_s(X_i)\right)_{i=1,\dots,n} + \varepsilon = \bar{Y} + \varepsilon \quad (2)$$

Recursive partitioning is well adapted to the problem we consider as it is particularly efficient in capturing the effect of thresholds (e.g., a low value for one parameter and a high value for another parameter, or when a certain parameter reaches a threshold value). One of the limitation of this regression is that it sometimes consider so many potential relations that it may over-fit.

4.1.4 Multivariate Adaptive Regression Splines (MARS)

MARS is a combination of (linear) spline regression, stepwise model fitting and recursive partitioning. A regression with a single input starts with a mean-only model and adds basis functions in a stepwise manner while adding the overall linear trend first. A second model using linear regression via least squares is fit to the data. This model is then added to the basis functions in a way that reduces the sum of squared error (SSE) between the observations and predictions. A fourth basis function is then added to minimize the SSE again. This process is repeated until M basis functions have been added. At this point, the MARS procedure will try to simplify the model using stepwise deletion of basis functions while keeping the y-intercept and linear trend. The $M-2$ candidate leading to the smallest increase of SSE will be selected. This deletion will be applied until regressed to the original linear model.
Stepwise addition and deletion leads to the creation of two sets of $M-2$ models. The "best" model is chosen using a generalized cross validation score which corresponds to a SSE normalized by the number of basis functions considered. With multiple inputs, the basis functions will consider main effects and multiple-way interactions. The options used for this analysis consider only two-way interactions to avoid the exponential cost of considering more interactions.
MARS usually leads to similar results as linear regression with a greater accuracy, and with the inclusion of non-monotonic effects and conjoint influences. However, it performs poorly with discrete inputs due to the use of splines.

4.2. Estimating global ranking using results from multiple regressions

A consequence of the use of multiple regressions is that the ranking of the inputs amongst themselves is not obvious when the different regressions disagree. A qualitative approach has been used in the past [7] based on the physics considered in the problem and expert knowledge, but such an approach introduced some subjectivity and is hard to document. It highlighted the need for a more quantitative approach.

Two effects of the uncertainty in the input on the output of consideration are estimated in the present study.

- The individual or "main" effect represents the influence of the uncertain input i by itself and is estimated with $R^2_{cont}(i)$ in the stepwise regression and S_i when response surfaces are considered

- The effect of the uncertain input from its interaction with other variables, which is not captured by the stepwise regression as it is an additive regression, and estimated with $(T_i - S_i)$ by the response surfaces.

The main effect of the uncertainty is usually the most important effect and the most stable estimate. As a result, the input variables are ranked according to their main global effect, estimated as follow:

$$M_i = \frac{R^2_{cont} + \sum_{j=1}^{n} R_j^2 . S_{i,j}}{n+1} \quad (3)$$

Where $R^2_{cont}(i)$is the influence of input i according to the linear or rank regression,R_j^2 the coefficient of determination one response surface j and $S_{i,j}$ the first order sensitivity index for input i when using response surface j. The cut-off for main contribution effect is set at 0.02, meaning that any main contribution lower than 0.02 is considered to be negligible.

The conjoint influence is estimated as an average between all the response surfaces considered

$$C_i = \frac{\sum_{j=1}^{n} R_j^2 . \left(T_{i,j} - S_{i,j}\right)}{n} \quad (4)$$

Due to the fact that conjoint influence is a more volatile indicator, the cut-off for influential variables is set to 0.1 (meaning 10 percent of the regression is explained via conjoint influence with this input). Furthermore, a variable is not considered as influent if its only influence is conjoint (no main effect) unless supported by the equations and what is known about the model.

4.3. Scatterplots

The use of scatterplots while qualitative, is yet a powerful technique that complement the suite of regression techniques as it allows to visually assess the relations between the inputs and outputs considered. It is used as a graphical confirmation that the relation estimated by any of the regression techniques is indeed present and not spurious.

5 EXAMPLE

In order to illustrate the methodology, we use the code to emulate a generic small diameter nozzle pipe with a reasonably large number of uncertain inputs (152) and 2 outputs. The sample size is set to 1,000.

5.1. Case set up

All the potential inputs are listed below (with the spatially varying inputs coloured in green). The unit when applicable as well as the codename used in the regression is added to each considered uncertain input.

- Effective Full Power Year (yr) **[EFPY]**
- Pipe wall thickness (m) **[THICK]**
- PWSCC initial flaw length (m)**[INILA##] [INILC##]**
- PWSCC initial flaw depth (m)**[INIDA##] [INIDC##]**
- Operating temperature (°C) **[TEMP]**
- Hoop WRS pre-mitigation (MPa) **[WRSAX_##]**
- Axial WRS pre-mitigation (MPa) **[WRSHP_##]**
- Surface-crack dist. Rule modifier **[SURFRUL]**

- TW crack dist. Rule modifier (mm) **[THWLRUL]**
- yield strength left-pipe (affects COD) (MPa) **[YS_LP]**
- ultimate strength left-pipe (affects COD) (MPa) **[UTS_LP]**
- elasticity modulus left-pipe (affects COD) (MPa) **[E_LP]**
- yield strength right-pipe (affects COD) (MPa) **[YS_RP]**
- ultimate strength right-pipe (affects COD) (MPa) **[UTS_RP]**
- elasticity modulus right-pipe (affects COD) (MPa) **[E_RP]**
- Proportionality constant A – model parameter for crack initiation (y^{-1} MPa^{-1}) **[A_AC_##] [A_CC_##]**
- Multiplier to the proportionality constant A – model parameter for crack initiation **[A_MULT]**
- Activation energy for crack growth (kJ/mol) **[QG]**
- Comp to Comp variability factor - model parameter for crack growth **[FCOMP]**
- within comp variability factor – model parameter for crack growth **[FLAWA_##] [FLAWC_##]**
- peak to valley ratio **[P2V]**
- characteristic width (mV) **[CHARWD]**

Each of the two outputs selected for this example represents one type of result that may be encountered in the analyses under consideration. The outputs are the following:

- **Probability of circumferential crack initiation over 60 yr**. which should be around 10^{-3} for 60 years. This is considered as a good test example with an output that occurs rarely considering the sample size of 1,000. In the example the number of occurrences of such event was 2 out of 1,000 realizations.
- **Number of axial cracks over 60 yr.** which remains a discrete function but not limited to two values and can vary between 0 and 19.

5.2. Results of regression analyses

The analyses have been performed twice. The first time using all the 152 inputs, and a second time with only the inputs that could impact this particular output (for instance, the probability of 1st circumferential crack or the number of axial crack occurring is not affected by any of the crack growth parameters). The purpose is to confirm that the method proposed is stable enough and the results are not significantly affected when applied on a large number of inputs.

5.2.1 1st Circumferential crack occurrence over 60 years

Table 1 shows the results of the regressions on the 1st circumferential crack occurrence. Despite the small number of occurrences, the component of the proportionality constant associated with the first crack is considered the most important parameter by all three regressions as expected from the crack initiation equations. All the other inputs are at noise level. The result is surprisingly good (especially for recursive partitioning) considering the likelihood of spurious correlation. It is also important to note that this is a case where rank regression may perform not as good as linear regression as the amplitude of the sampled value for the proportionality constant A is extremely important.

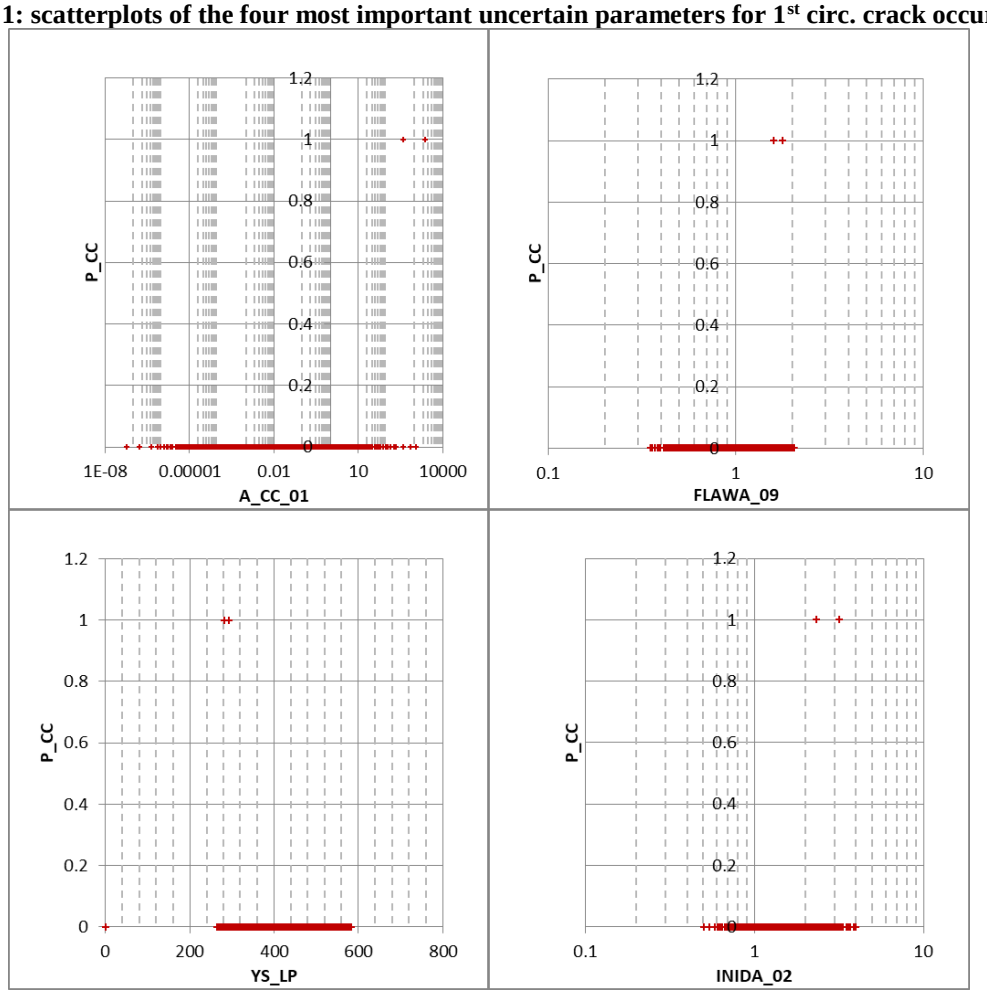
Table 1: Regressions analyses results on 1st circ. crack occurrence

	Rank Regression			Recursive Partitioning		MARS		Main Contri-bution	Conjoint Contri-bution *
Final R^2	0.06			0.4		1			
Input	R^2 inc.	R^2 cont.	SRRC	S_i	T_i	S_i	T_i		
A_CC_01	0.01	0.01	0.01	1.00	1.00	0.01	1.00	0.14	0.50
INIDA_01	0.01	0.01	-0.01	---	---	0.00	0.77	0.00	0.38
FLAWA_09	0.02	0.01	0.01	---	---	0.00	0.52	0.00	0.26
YS_LP	0.03	0.01	-0.01	---	---	0.00	0.82	0.00	0.41
INIDA_02	0.03	0.01	0.01	---	---	0.00	0.85	0.00	0.43
WRSAX_20	0.02	0.01	-0.01	---	---	0.00	0.41	0.00	0.20
A_AC_16	0.04	0.01	-0.01	---	---	0.00	0.87	0.00	0.43
FLAWA_10	0.04	0.00	0.01	---	---	---	---	0.00	0.00
FLAWA_14	0.05	0.00	0.01	---	---	---	---	0.00	0.00
INILA_10	0.05	0.00	0.00	---	---	---	---	0.00	0.00
INIDA_14	0.05	0.00	0.00	---	---	---	---	0.00	0.00
INIDC_02	0.05	0.00	0.00	---	---	---	---	0.00	0.00
A_AC_01	0.06	0.00	0.00	---	---	---	---	0.00	0.00
A_AC_17	0.06	0.00	0.00	---	---	---	---	0.00	0.00
WRSAX_12	---	---	---	---	---	0.00	0.55	0.00	0.27
WRSAX_21	---	---	---	---	---	0.00	0.00	0.00	0.00
WRSAX_17	---	---	---	---	---	0.00	0.83	0.00	0.42
WRSAX_10	---	---	---	---	---	0.00	0.21	0.00	0.10

** highlighted in yellow if conjoint contribution is larger than 0.1*

The set of scatterplots (Figure 1) for the four identified most important inputs shows the issue when a really small number of realizations occurs. Some of the 152 inputs may have high values (or low values) associated with the 2 realizations leading to circ. crack initiation and therefore be considered as important.

Figure 1: scatterplots of the four most important uncertain parameters for 1st circ. crack occurrence



The reduced analysis considered only 7 inputs parameters (EFPY, THICK, TEMP, A_MULT, WRSAX01, A_CC_01, A_CC_02). The results of the regression are displayed in Table 2. They are similar to the larger regression analysis, essentially identifying one important parameter. Such results demonstrate that while adding more input tends to make the analysis noisier, it does not seem to affect the regressions values significantly to the point of changing any conclusion.

Table 2: Regressions analyses results on 1st circ. crack occurrence with reduced number of inputs

	Rank Regression			Recursive Partitioning		MARS		Main Contri-bution	Conjoint Contri-bution *
Final R^2	0.02			0.4		1			
Input	R^2 inc.	R^2 cont.	SRRC	S_i	T_i	S_i	T_i		
A_CC_01	0.01	0.01	0.01	1.00	1.00	0.01	1.00	0.14	0.50
WRSAX_01	0.01	0.00	0.01	---	---	0.00	0.00	0.00	0.00
TEMP	0.02	0.00	0.00	---	---	0.00	0.00	0.00	0.00
A_MULT	0.01	0.01	0.01	---	---	0.00	0.12	0.00	0.06
EFPY	---	---	---	---	---	0.00	0.01	0.00	0.00
THICK	---	---	---	---	---	0.00	0.95	0.00	0.47
A_CC_02	---	---	---	---	---	0.00	0.00	0.00	0.00

** highlighted in yellow if conjoint contribution is larger than 0.1*

1.1 Number of axial cracks occurring over 60 years

On an output with a greater likelihood of occurrences, two of the three regression analyses identify the same more important parameter. MARS tends to mostly capture the conjoint influence which is consistent with the equation estimating crack initiation.

As expected the components of the A parameter play a significant role both as a multiplier and individual values (Table 3). The different location of WRS appear several time. It is not unexpected considering that the values at different locations are all correlated in order to smooth WRS profiles.

Table 3: Regressions analyses results on number of axial cracks (Rep #1)

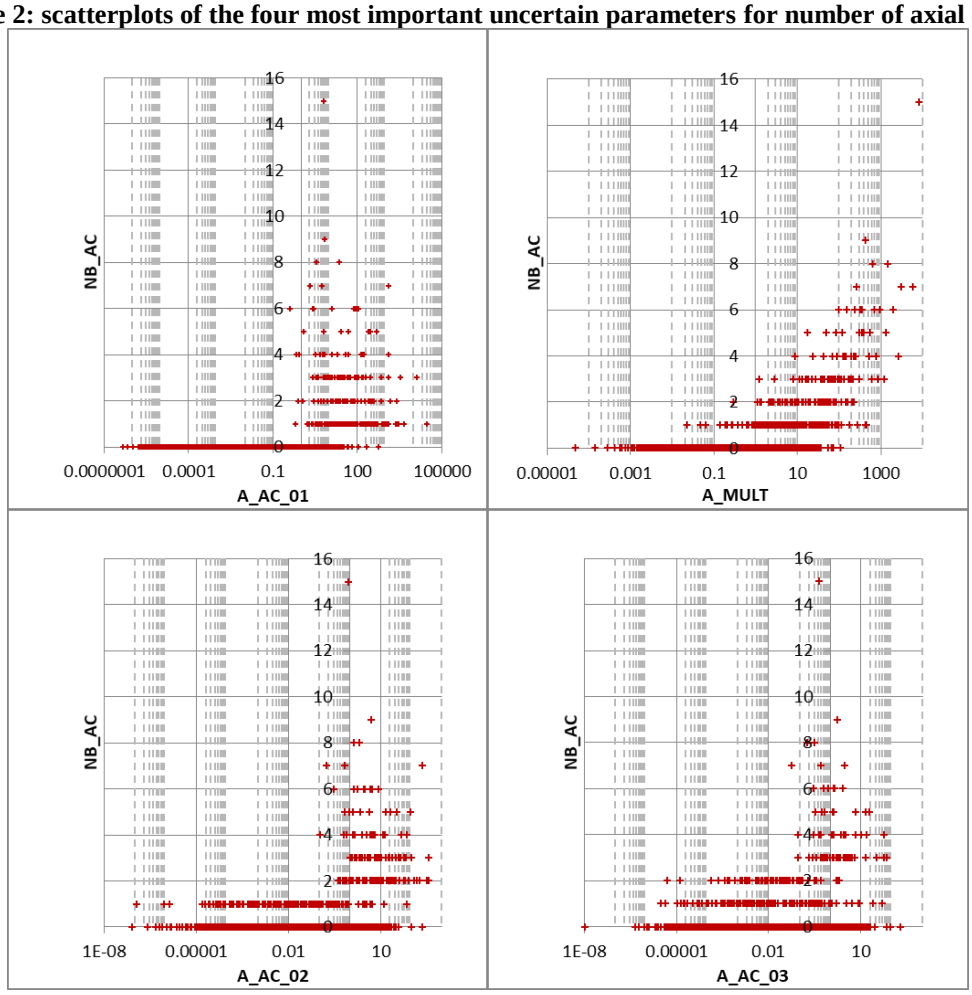
	Rank Regression			Recursive Partitioning		MARS		Main Contri-bution	Conjoint Contri-bution *
Final R^2	0.71			0.94		0.9			
Input	R^2 inc.	R^2 cont.	SRRC	S_i	T_i	S_i	T_i		
A_AC_01	0.52	0.52	0.40	0.13	0.23	0.00	0.45	0.22	0.25
A_MULT	0.66	0.14	0.29	0.37	0.78	0.00	1.00	0.16	0.64
A_AC_02	0.69	0.03	0.12	0.03	0.24	0.00	0.00	0.02	0.10
A_AC_03	0.69	0.01	0.06	0.01	0.13	0.02	0.98	0.01	0.49
WRSHP_13	0.70	0.01	0.04	---	---	0.00	0.00	0.00	0.00
WRSHP_08	---	---	---	0.00	0.01	---	---	0.00	0.00
INILC_02	0.70	0.00	-0.04	---	---	---	---	0.00	0.00
WRSHP_15	---	---	---	0.00	0.10	0.00	0.57	0.00	0.30
WRSHP_23	---	---	---	---	---	0.00	0.16	0.00	0.07
INIDA_14	0.71	0.00	-0.04	0.00	0.00	---	---	0.00	0.00
A_AC_13	---	---	---	0.00	0.00	---	---	0.00	0.00
TEMP	0.70	0.00	0.04	---	---	0.00	0.00	0.00	0.00
WRSHP_10	0.71	0.00	0.03	---	---	0.00	0.00	0.00	0.00
WRSHP_02	---	---	---	0.00	0.01	---	---	0.00	0.01
WRSHP_24	---	---	---	---	---	0.00	0.00	0.00	0.00

** highlighted in yellow if conjoint contribution is larger than 0.1*

Scatterplots are displayed in Figure 2. The influence of the components of the A parameters are as expected: A_{mult} affects all crack initiation so the higher its value is, the more likely it is to have multiple cracks.

The first location component of A_{01} affects the first crack occurrence but does not distinguishes between 1 and multiple cracks (the bottom line is not affected by the values, all other lines are). Similarly, A_{02} affects the condition for having two cracks (the two bottom lines are not affected by its values, all lines above are) and A_{03} (the three bottom lines are not affected by its values, all lines above are).

Figure 2: scatterplots of the four most important uncertain parameters for number of axial cracks



The reduced analysis is performed using 24 inputs (EFPY, THICK, TEMP, A_MULT, WRSHP01, A_AC_01 to A_AC_19) and the corresponding results are displayed below (Table 4). The results are consistent with the non-reduced analysis

Table 4: Regressions analyses results on number of axial cracks with reduced number of inputs

	Rank Regression			Recursive Partitioning		MARS		Main Contri-bution	Conjoint Contri-bution *
Final R^2	0.71			0.94		0.89			
Input	R^2 inc.	R^2 cont.	SRRC	S_i	T_i	S_i	T_i		
A_AC_01	0.52	0.52	0.41	0.17	0.27	0.00	0.75	0.23	0.38
A_MULT	0.66	0.14	0.29	0.36	0.73	0.05	1.00	0.17	0.60
A_AC_02	0.69	0.03	0.12	0.03	0.21	0.00	0.61	0.02	0.35
A_AC_03	0.69	0.01	0.06	0.00	0.05	0.00	1.00	0.00	0.47
A_AC_04	0.70	0.00	0.03	0.01	0.23	0.00	0.00	0.00	0.10
TEMP	0.70	0.00	0.04	---	---	0.00	0.72	0.00	0.32
A_AC_07	0.70	0.00	-0.02	0.00	0.02	0.00	0.61	0.00	0.28
A_AC_13	---	---	---	0.00	0.00	---	---	0.00	0.00
A_AC_08	---	---	---	---	---	0.00	0.30	0.00	0.13
A_AC_19	0.70	0.00	-0.02	---	---	---	---	0.00	0.00
A_AC_09	0.71	0.00	-0.02	---	---	0.00	0.00	0.00	0.00
WRSHP_01	0.70	0.00	0.05	0.00	0.01	0.00	0.00	0.00	0.00
THICK	0.70	0.00	-0.02	---	---	0.00	0.00	0.00	0.00

** highlighted in yellow if conjoint contribution is larger than 0.1*

6 CONCLUSION

These illustrative examples highlight several aspects of the propose approach that can be summarized with the following key points:

- Even with a very small number of occurrences, linear or rank regression and recursive partitioning can identify the most important parameters that can be considered for importance sampling.
- MARS works better when the output is continuous or with enough discrete values.
- The regressions methods reach satisfying level of stability even when the number of events of interest is small (less than 10)
- When many inputs variables are strongly correlated such as in WRS, considering only one representative value leads to cleaner analyses.
- While it is recommended to reduce the number of inputs to those which have really an impact to avoid spurious correlation (especially when few events occur), the regressions analyses still identify the most important variables when the input set include a reasonably large number of non-influential variables (more than 130 in the case of probability of circ. crack and number of axial cracks).

In addition a more complete analysis considering replicated Monte Carlo and larger sample size has confirmed that the most important inputs have been correctly identified.
Beyond the application of this approach to a larger number of test problem and real-life problem, we plan to test other response surfaces and regressions techniques that may be more appropriate depending on the problem and output under consideration.

Acknowledgements
The development of the Version 2.0 xLPR code is a group effort that spans a variety of experts across many fields of expertise from the U.S. NRC, EPRI, and their contractors. The ongoing success of this program emphasizes the dedication of the xLPR team, the strength of its leadership, and the support from both the NRC and EPRI.

7 REFERENCES

[1] D. Rudland and C. Harrington, "xLPR Pilot Study Report," U.S NRC-RES Washington, DC and EPRI, Palo Alto, CA, NUREG-2110 and EPRI 1022860, May 2012.

[2] D. Rudland, C. Harrington and R. Dingreville, "Development of the extremely low probability of rupture (xLPR) Version 2.0 code," in *Proceedings of the ASME 2015 Pressure Vessels & Piping Division*, Boston, MA, USA, 2015.

[3] J. C. Helton and F. J. Davis, "Latin Hypercube Sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering and System Safety,* vol. 81, no. 1, pp. 23-69, 2003.

[4] C. B. Storlie and J. C. Helton, "Multiple predictor smoothing methods for sensitivity analysis: Description of Techniques," *Reliability Engineering and System Safety,* vol. 93, no. 1, pp. 28-54, 2008.

[5] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana and S. Tarantola, Global Sensitivity Analysis - The primer, ISBN 978-0-470-05997-5: Wiley, 2008.

[6] A. Saltelli, K. Chan and E. M. Scott, Sensitivity Analysis, Wiley series in probability and statistics, 2000.

[7] Sandia National Laboratories, "State-of-the-Art Reactor Consequence Analyses Project - Uncertainty Analysis of the Unmitigated Long-Term Station Blackout of the Peach Bottom Atomic Power Station," for the Office of Nuclear Regulatory Research - U.S. Nuclear Regulatory Commission, NUREG/CR-7155 SAND2012-10702P, 2012.

[8] I. Sobol, "Sensitivity estimates for nonlinear mathematical models," *Mathematical Modeling & Computational Experiment,* vol. 1, pp. 407-414, 1993.