INL/EXT-11-22936

Uncertainty Analysis for RELAP5-3D

Aaron J. Pawel
George L. Mesina

August 2011

Idaho National Laboratory

The INL is a U.S. Department of Energy National Laboratory
operated by Battelle Energy Alliance

INL/EXT-11-22936

Uncertainty Analysis for RELAP5-3D

Aaron J. Pawel
George L. Mesina

August 2011

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Office of Nuclear Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

UNCERTAINTY ANALYSIS FOR RELAP5-3D

ABSTRACT

In its current state, RELAP5-3D is a "best-estimate" code; it is one of our most reliable programs for modeling what occurs within reactor systems in transients from given initial conditions. This code, however, remains an estimator. A statistical analysis has been performed that begins to lay the foundation for a full uncertainty analysis. By varying the inputs over assumed probability density functions, the output parameters were shown to vary. Using such statistical tools as means, variances, and tolerance intervals, a picture of how uncertain the results are based on the uncertainty of the inputs has been obtained.

INTRODUCTION

The Reactor Excursion and Leak Analysis Program (RELAP5) that is developed by scientists at Idaho National Laboratory (INL) is the most popular program in the world for analyzing transients in reactor systems. However, this program is known as a "best-estimate" code. The output from RELAP5 is not exact; it is an approximation of the truth [1].

The fact that the code is an approximation should be expected. Computer programs are incapable of performing exact calculations at the level of mathematics that is required to solve reactor transients. Basic arithmetic operations may be performed with a reasonable level of accuracy, although floating-point arithmetic is subject to round-off error. Such things as the calculation of a derivative function or other non-discrete mathematics, however, are beyond exact calculation for a machine. Therefore, when such operations are required, the programmer instructs the program to perform the operation by various techniques for numerical evaluation of the desired result. This sort of approximation introduces errors into the scheme [4].

Additionally, computers are finite machines. Therefore, no irrational result and few rational results may be exactly represented within the computer; instead, these values are represented by the nearest value that may be obtained with the bits available to the machine. This approximation is carried through all subsequent evaluations and the error tends to increase rather than decrease, although systems exist where the error does not grow [4].

Another source of error in the algorithms employed by computer codes that model reactor transients of this sort is that the physics of two-phase flow is not well-understood. Moreover, many correlations have been developed for certain ranges of parameters, but these are often used outside of their range of validation. Such tactics introduce more error into the system [4].

There are more sources of error within computer codes than are listed here, but they have impact nonetheless. Therefore, a computer code's evaluation is not complete until an uncertainty analysis has been performed on the program [4].

Uncertainty analysis may be performed in many ways. The approach detailed herein involves propagating uncertainties through the code and basing the decision on the applicability of the code on the statistical evaluation of the outcome. Statistical methods that are important for the kind of uncertainty analysis performed here include means, variances, standard deviations, order statistics, and tolerance intervals. All of these statistical evaluators, with the exception of a tolerance interval, may be applied to both samples and populations. The tolerance interval should be applied to samples and a confidence interval applied to populations [2].

MATERIALS AND METHODS

Expert Consultation

Test cases for this experiment were chosen from a set of Separate Effects Tests and Integral Effects Tests that make up the RELAP5-3D Developmental Assessment [3]. These experiments are large, and often international, experiments that the code is validated against. The investigative team chose to begin implementation of this analysis on FLECHT-SEASET Test 31701 and to continue the analysis with Marviken Critical Flow Test 22.

Once the test case had been identified, the major output of that test was identified as the Figure Of Merit (FOM). Peak cladding temperature was selected to be the FOM for the FLECHT-SEASET case and the maximum of the time-dependent critical flow rate was selected as the FOM for the Marviken case. The Marviken selection is deemed reasonable because it is believed that the uncertainty in the flow rate at any point during the transient is bounded by the uncertainty at the maximum.

With the FOM known, copious input information must be obtained by varying important input parameters through a range of reasonable values in order to perform a statistical analysis. An INL expert was consulted for the identification of the input parameters that were important for the determination of the FOM for the FLECHT-SEASET case. For the Marviken experiment, these parameters were selected from among those that were different between Critical Flow Test 22 and a similar experiment.

After identifying the significant input parameters, experts were consulted to obtain a realistic range for each of these parameters. By deciding on a plausible variation about the nominal value for each of these input parameters, the expert provides a reasonable uncertainty band on the input. The selected parameters for the two tests and their ranges are shown in Table 1.

Table 1. The parameters and relevant ranges chosen for variation for the two tests.

	FLECHT-SEASET Test 31701		Marviken Critical Flow Test 22	
Parameter 1	System Pressure	40 ± 10 psi	Pressure Vessel Temp.	484 ± 0.6 K
Parameter 2	Inlet Temperature	$127 \pm 4°\mathrm{F}$	Discharge Pipe Temp.	441 ± 0.6 K
Parameter 3	Reflood Rate	$6.1\ ^{\mathrm{in}}/_{\mathrm{sec}} \pm 10\%$	System Pressure	4930 ± 9 kPa
Parameter 4	Peak Power	$2.3\ ^{\mathrm{kW}}/_{\mathrm{m}} \pm 10\%$	Test Nozzle Diameter	0.5 m $\pm\ 0.1\%$

Automation

This uncertainty is to be propagated through the execution of the code for the uncertainty analysis. However, to do so could take copious amounts of time. Instead, the process was automated with a C-shell script. A matrix of parameter values was hardcoded into the script, which then ran RELAP5-3D in a set of nested loops.

The input decks for each run must be different because RELAP is a deterministic code. Therefore, a FORTRAN program was written to modify an existing input deck. The program copies an existing input deck to a new file, modifying the lines as instructed by well-placed comment cards and the parameters selected by the nested loops of the script. The new input deck is written to a file that has the indices of the selected input parameters attached to it to create a unique file name. This program is called in the nested loops prior to RELAP.

The FOM must also be gathered from the output file. This process is also automated by a FORTRAN program. A special control variable was added to the original input deck that would track the chosen FOM as RELAP moved through its time steps. This way, the program merely needs to scan the output file for the character string that is represents the FOM and collect the number. This program then writes the input parameters and the FOM out to a new statistics file that is also marked by the indices of the input parameters that produced this output.

The four parameters in Table 1 were varied among 9 values each in their own ranges. To collect the relevant statistics for all possible combinations of these parameters requires $9^4 = 6561$ runs of RELAP5-3D. Each run of this job takes nearly 10 seconds. The total sequential runtime on a single computer core requires about 65610 seconds or 18.2 hours. This is too time-consuming for the purposes of repeating the job to study different choices of parameters or parameter values, for example varying the values according to a Uniform or Normal distribution (Equations 1 and 2 below). Moreover, collecting population statistics from a set with much larger sets of parameter values, as is routinely done in industry, was unworkable in the timeframe available.

The original job-running script was therefore rewritten in Python to enable parallel operations on an INL cluster named Fission. With the help of INL's High Performance Computing staff, the script enabled each pass through the nested loops to run on a different core. The 18.2 hour job of running 6561 10-second RELAP5-3D runs was reduced to about twenty minutes of wallclock time. Additionally, when the size of the population was upgraded from nine values per parameter to sixty-three, the 15.7 million transients required less than a day.

Statistical Distributions

Uniform distributions involve evenly distributing the probability over all values that the distribution could take on. In essence, the probability of the random variable taking on a value, $x \in [a, b]$, is

$$P(x) = \frac{1}{b-a} \tag{1}$$

In this way, the area under this horizontal line equates to the unity and the laws of probability are satisfied.

The other statistical distribution that is important in this kind of work is the Gaussian distribution, which is also known as a normal. In this distribution, the probability that an observation returns a given $x \in (-\infty, \infty)$ is

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2}$$

The Central Limit Theorem indicates that the distribution of statistical parameters such as the mean approaches a Gaussian as the number of observations becomes infinite.

The selected parameters were varied through a uniform distribution. This choice was made both for ease of variation and for statistical validity. Parameters may easily be varied through a uniform distribution because each value need only occur once in the loop, and the values may be evenly spaced through the interval chosen for them. This distribution choice is safe because due to the Central Limit Theorem, the results are expected to either take on a uniform or a normal distribution. These distributions are easy to evaluate based on their probability functions.

Statistics Equations

Statistics textbooks everywhere make reference to such statistical values as described above. The mean (μ) is the expected value of the FOM. The sample mean ($\hat{\mu}$) is the arithmetic average of the sample values of the FOM that are calculated by RELAP5-3D.

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3}$$

x_i is the i^{th} value in the sample and n is the number of values within the sample.

A sample's variance is a measure of the degree of separation of the sample values around the mean. A small variance indicates that the sample values are packed tightly around the mean. The variance, σ^2, is estimated by the sample variance given in Equation (4).

$$\hat{\sigma}^2 = \frac{\sum_i (x_i - \hat{\mu})^2}{n-1} \tag{4}$$

where x_i is the i^{th} value in the sample, n is the number of values within the sample, and $\hat{\mu}$ is the sample mean. The standard deviation is the square root of the variance.

Order statistics involve ranking the values in the sample from maximum to minimum. A percentile rank is given to a point by determining what percentage of the points fall below that point. Order statistics help to give one a feel for the way that the outcomes are distributed. If it is necessary for a parameter to be less than a certain (e.g. regulatory) limit, one may use the percentile rank of that limit to determine how likely it is that the value falls below that number.

One may also use a one-sided tolerance interval to determine this likelihood. The tolerance interval that informs the interpreter that a fraction of the population, α, is within the specified tolerance limit with a probability β is called an α/β tolerance interval. A two-sided tolerance interval gives both an upper and lower bound on this interval, while a one-sided tolerance interval gives only one of the two. For a one-sided tolerance interval, the missing limit is the corresponding infinity.

For reduction of sample size, the size of the selection is influenced by Wilks' theorem for tolerance intervals. A one-sided α/β tolerance interval needs at least a sample size of [2]

$$n \geq \frac{\ln(1-\beta)}{\ln \alpha} \tag{5}$$

Wilks' 1941 work produces a formula to be used to determine the minimum sample size [2]:

$$n\alpha^{n-1} - (n-1)\alpha^{n} = 1 - \beta \tag{6}$$

When this formula is solved recursively, the minimum size of a sample is obtained. Some common tolerance intervals and the number of observations, *n*, necessary to create them are shown in Table 2 below.

Table 2. Statistical Tolerance Limits and their minimum sample sizes. Adapted from Lavin [2].

	One-sided Tolerance Interval			Two-sided Tolerance Interval		
β \ **α**	**0.90**	**0.95**	**0.99**	**0.90**	**0.95**	**0.99**
0.90	22	45	230	38	77	388
0.95	29	59	299	46	93	473
0.99	44	90	459	64	130	662

Sample Size Reduction

It is often impractical to deal with the large number of observations required for gathering good statistical information. Therefore, various methods of sample size reduction are employed. Two popular methods were chosen for use in the uncertainty analysis performed. The Latin Hypercube (LH) and Stratified Sampling (SS) were chosen because one does not need to know anything about the model being evaluated in order to conduct the uncertainty analysis [2].

In LH sampling, n-dimensional space is constructed, and each of the n parameters has its m values placed along an axis. One value is selected from each axis, and an n-tuple is created with these values. The process is repeated by selecting a different value from each axis and forming another n-tuple until m points have been formed. The sample consisting of these m points is representative of the population.

In stratified sampling, the sample space is broken up into regions called strata. A sampling technique is applied to the strata. This sampling technique is often either random sampling or a systematic sampling designed to reduce the sampling error. The resulting sample is representative of the population. INL statisticians recommend that multiple points be sampled from each strata so that a within-stratum variance may be obtained and standard error may be calculated.

Prior to the re-write of the processing script, the population being considered consisted only of nine values from each of the four parameters. This set would only allow for a LH of nine points, which is below the threshold given in Table 2 for a 95/95 one-sided tolerance interval.

An interesting method for attaining the 95/95 level with just nine parameter values via a different type of hypercube was devised. For even numbers of parameters, ordered pairs may be placed along each axis. In this way, the number of points increased from nine to eighty-one (surpassing the threshold of 59 from Table 2). One advantage of this approach is that, unlike the four-dimensional hypercube, it can be visualized as shown in Figure 1.

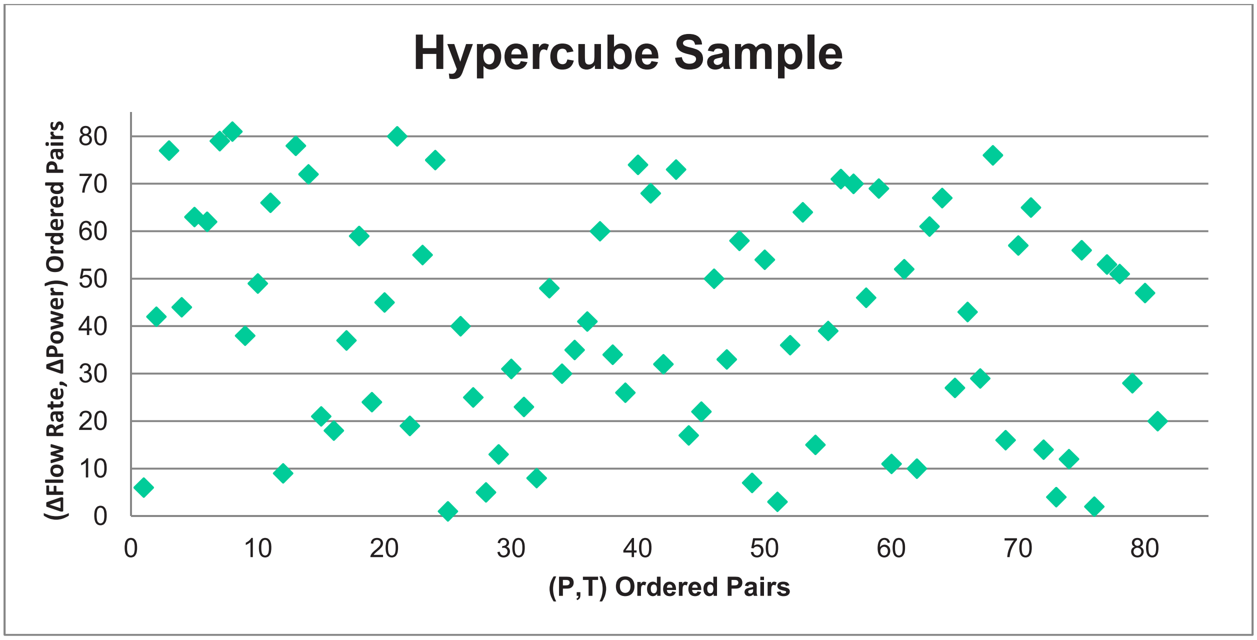

Figure 1. A hypercube with ordered pairs on each axis rather than individual values.

The parameters are numbered for low to high value. The ordered pair index is calculated by the formula 9×(index of parameter 1) + (index of parameter 2). The statistical results are comparable to eighty-one values on each of the four axes and are thus better than those from a nine-point hypercube. However, this selection of eighty-one 4-tuples is not as random as the typical set of LH 4-tuples would be. This will be the subject of future study.

RESULTS

FLECHT-SEASET Test 31701

The statistical evaluation of the FLECHT-SEASET results is shown in Table 3 below. The Latin Hypercube and Stratified Sample values are the averages of ten trials because every time the sample is retaken the values change.

Table 3. Relevant statistical parameters for the evaluation of the uncertainty in the peak cladding temperature of the core in FLECHT-SEASET Test 31701 as calculated by RELAP5-3D.

	Population	Latin Hypercube	Stratified Sample
μ	1159.44	1159.97	1159.96
σ^2	8.58	8.20	8.13
σ	2.93	2.86	2.85
σ (% of μ)	0.25	0.25	0.25
P_{95}	1164.40	1164.59	1164.51
Maximum	1168.82	1166.33	1166.51
1-Sided T.I.	1164.26	1164.69	1164.66

The means are all close together, with the largest percent difference of 0.05% existing between the LH sampling and the total population means. The Developmental Assessment largest measured value from the experiment is just below 1200 K, and no uncertainty band is given. Therefore, it is difficult to assess how the means presented here match up to the measured data other than to observe that the Developmental Assessment passes the code for how well it models the actual facility [3].

The standard deviations are all small, at about one-quarter of a percent of the mean. These small standard deviations suggest that for reasonable uncertainties in the input information, RELAP5-3D is sure of its output. The probability distribution of the peak cladding temperature for the evaluation of FLECHT-SEASET Test 31701 is shown in Figure 2 below. It resembles a normal distribution, save for the small depression around 1160 K.

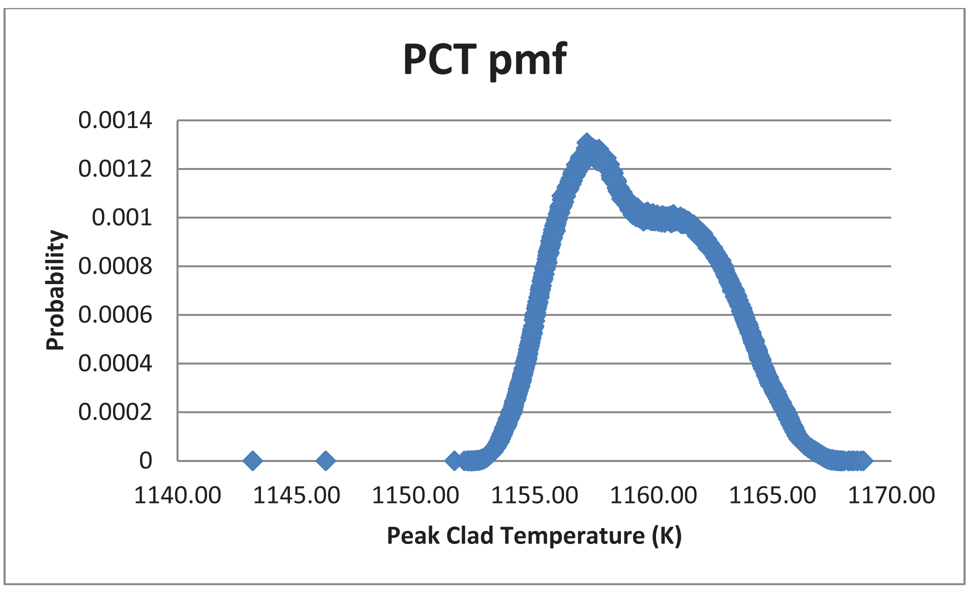

Figure 2. The probability distribution for the FOM of the FLECHT-SEASET Test 31701.

These results help validate the statistical calculations because the calculated means fall near the center of the hump, near the beginning of the plateau. The curve is narrow, encompassing less than twenty kelvins, so the standard deviation should be very small. Lastly, the ninety-fifth percentile looks like it should occur near 1165 K, and Table 3 gives its value as 1164.40 K.

Marviken Critical Flow Test 22

The Marviken Critical Flow Test 22 statistical results are shown in Table 4 below. The Latin Hypercube and Stratified Sample values are the averages of ten trials since every time one of these reduced-sample techniques is used, the values change.

Table 4. The relevant statistical parameters of the analysis of the maximum critical flow rate in the modeling of Marviken Critical Flow Test 22.

	Population	Latin Hypercube	Stratified Sample
μ (kg/s)	15092.30	15091.31	15092.78
σ^2 (kg^2/s^2)	24214.60	23372.51	24111.34
σ (kg/s)	155.61	152.82	155.27
σ (% of μ)	1.03	1.01	1.03
P_{95} (kg/s)	15361.95	15346.95	15359.65
Maximum (kg/s)	15416.99	15394.75	15410.55
1-Sided T.I. (kg/s)	15348.28	15343.70	15348.99

The largest percent difference in the means is again between the LH and the population means, at 0.007%. The measured value shown in the Developmental Assessment is approximately 13000 kg/s with an uncertainty of 15%. These values are on the edge of the uncertainty band, but have been accepted by the Developmental Assessment [3].

The standard deviations are large for this model, but they remain approximately one percent of the mean, which again suggests that RELAP5-3D is as certain of its output as the user is of its input.

It is believed, however, that errors have been made in the input-processing of these runs, likely in the python script that runs the code. The probability function shown below is the result of the calculations, and is what informed the investigative team that something is wrong. The input processing has been slightly improved by tweaking the naming algorithm, but there remains an irregular tail on the function which is as of yet unexplained.

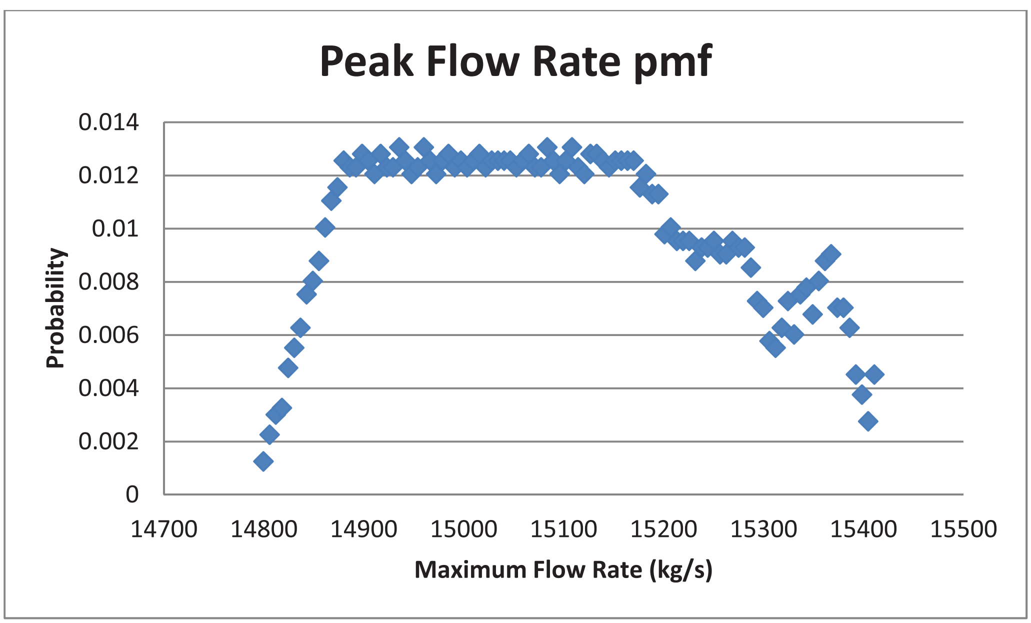

Figure 3. Probability function for the FOM of Marviken Critical Flow Test 22.

FUTURE WORK

The theoretical novelty of using ordered pairs in a Latin Hypercube will be compared with the normal Latin Hypercube, wherein the ordered 4-tuples are completely random, to determine the extent to which the latter is statistically better, for example less biased. The extent of improvement will be determined. If there is no significant difference, this will be useful in reducing the number of parameter values needed in Latin Hypercube sampling.

The most pressing future work is to determine what is going wrong in the python script and why the probability function shown in Figure 3 looks the way that it does. This process will involve dissecting the script and examining the nested loops closely.

The ultimate goal of this work is to give RELAP5-3D the ability to perform its own uncertainty analysis. This capability requires performing this sort of analysis on the remaining models from the Developmental Assessment. Then, the code must be adapted to perform a concise, but accurate, uncertainty analysis on each transient that it runs, which will involve the creation of new subroutines in the RELAP5-3D code architecture.

ACKNOWLEDGEMENTS

This work was made possible by generous support from the Science Undergraduate Laboratory Internship (SULI) program, funded by the Department of Energy's Office of Science, and the Nuclear Science and Engineering Division at Idaho National Laboratory. Resources and advice were given by many members of the RELAP5-3D Code Development Team, including Cliff Davis, Paul Bayless, and Nolan Anderson for which we are most grateful.

REFERENCES

[1] RELAP5-3D Code Development Team, “Volume I: Code Structure, System Models, and Solution Methods,” *RELAP5-3D Code Manual,* June, 2005.

[2] R.B. Lavín, "Uncertainty and Sensitivity Analyses: Basic Concepts," S.UN.COP 2011, April, 2011.

[3] RELAP5-3D Code Development Team, “Volume III: Developmental Assessment,” *RELAP5-3D Code Manual,* September, 2009.

[4] A. Petruzzi, "The Origins of the Uncertainty," S.UN.COP 2011, April, 2011.