

AN ASSESSMENT OF UNCERTAINTY ON A LOFT L2-5 LBLOCA PCT BASED ON THE ACE-RSM APPROACH: COMPLEMENTARY WORK FOR THE OECD BEMUSE PHASE-III PROGRAM

KWANG-IL AHN*, BUB-DONG CHUNG and JOHN C. LEE¹

Korea Atomic Energy Research Institute

Dukjin-dong 150, Yuseong-gu, Daejeon, 305-353, Korea

¹University of Michigan, 2355 Bonisteel Blvd., Ann Arbor, MI 48109-2104, USA

*Corresponding author. E-mail : kiahn@kaeri.re.kr

Received August 31, 2009

Accepted for Publication November 26, 2009

As pointed out in the OECD BEMUSE Program, when a high computation time is taken to obtain the relevant output values of a complex physical model (or code), the number of statistical samples that must be evaluated through it is a critical factor for the sampling-based uncertainty analysis. Two alternative methods have been utilized to avoid the problem associated with the size of these statistical samples: one is based on Wilks' formula, which is based on simple random sampling, and the other is based on the conventional nonlinear regression approach. While both approaches provide a useful means for drawing conclusions on the resultant uncertainty with a limited number of code runs, there are also some unique corresponding limitations. For example, a conclusion based on the Wilks' formula can be highly affected by the sampled values themselves, while the conventional regression approach requires an a priori estimate on the functional forms of a regression model. The main objective of this paper is to assess the feasibility of the ACE-RSM approach as a complementary method to the Wilks' formula and the conventional regression-based uncertainty analysis. This feasibility was assessed through a practical application of the ACE-RSM approach to the LOFT L2-5 LBLOCA PCT uncertainty analysis, which was implemented as a part of the OECD BEMUSE Phase III program.

KEYWORDS : Uncertainty Analysis, Optimal Algorithm, ACE-RSM, LBLOCA, PCT

1. INTRODUCTION

The OECD Best Estimate Methods-Uncertainty and Sensitivity Evaluation (BEMUSE) Programme [1,2] has been performed: (1) to evaluate the practicability, the quality, and the reliability of the Best-Estimate (BE) methods, including uncertainty evaluation in applications relevant to nuclear reactor safety and (2) to promote the use of BE-methods within the regulatory bodies and the industry. When applying these methods to a transient of type Large-Break Loss of Coolant Accident (LBLOCA), the BEMUSE activity consists of two main steps: (1) Best-estimate and uncertainty and sensitivity evaluations of the LOFT L2-5 test (Phases 2 [1] and 3 [2]), with an a priori knowledge of the uncertainty methodologies presented in Phase I. The LOFT experiment L2-5 experiment was configured to simulate a double-ended 200% cold leg break (LBLOCA); (2) *Step 2*: Best-estimate and uncertainty and sensitivity evaluations of a nuclear power plant (Phases 4-5).

One of the key issues of the BEMUSE Phases 2 & 3 is to determine the appropriate number of statistical samples that must be evaluated for the PCT uncertainty. More specifically, one of the observations at the end of BEMUSE Phase 2 was that the 59-run utilized as a reference case (i.e., the first order of Wilks' formula [3,4]) was found to be too small to obtain a robust conclusion in the PCT statistics. Thus, the number of code runs needed to be increased from over 150 to 200 code runs (i.e., Wilks' formula at the order 4 or 5 for the estimation of the 95% percentile) in order to reduce the dispersion of the uncertainty in the prediction, and consequently, to obtain a more robust conclusion.

The Wilks' formula is a kind of statistically sound nonparametric order statistics, which provides a useful tool for estimating the relevant output uncertainty (mainly fractiles), especially when there is little information on the output PDF and a small sample size is available for code runs. For this approach, all uncertainty contributors

are sampled simultaneously in each run, and the number of code runs needed is only determined by both the requested coverage of the population (i.e., tolerance limit) and the confidence level. In real situations, however, the fractiles estimated by the approach also has a greater or lesser dependency on the size of samples because it employs random sampling schemes. As such, a concern often expressed about the sampling-based uncertainty analysis is how to determine the appropriate number of samples so that a statistical meaning could be obtained within a reasonable cost of the analysis.

From this point of view, the Response Surface Model (RSM) [5-7] has been used to give an appropriate surrogating model to an original model by using several kinds of experimental designs (such as fractional factorial designs or central composite designs) or a limited number of statistical samples (such as Latin Hypercube Samples) over all possible ranges of inputs, regression techniques, and optimization methods. This approach becomes very useful when it takes a long time for a code run to obtain the relevant output values (such as a physical trend of performance parameters in a complex physical model), and thus, the number of evaluations through such a model is limited to, at most, from tens to hundreds. The resultant model is a simple and high speed alternative model that can best describe the trend of the relevant output variables with the specified inputs. Since the functional form of the RSM is generally not known for many cases, it requires a tedious and time-consuming process to obtain the most relevant functional form between input and output parameters (coupled with multivariate nonlinear regression analysis).

The above limitation involved in the traditional RSM approach can be greatly reduced by employing the

Alternating Conditional Expectation (ACE) method [8-11]. The ACE method is a generalized regression algorithm that yields an optimal relationship between a dependent variable and multiple independent variables by maximizing the statistical correlation between the transformed dependent variable and the sum of transformed independent variables. This fundamental feature of the ACE algorithm requires an iterative convergence to the transformations for independent and dependent variables, while it does not require an a priori estimate of the functional forms of transformed variables. Once the optimal transformations are obtained, a simple regression analysis can be performed to determine the functional forms for the transformed dependent and independent variables. Thus, the ACE method offers two distinctive advantages over the traditional nonlinear regression analysis. First, while nonlinear regression analysis an a priori estimate of functional forms and sufficiently estimates accurately the fitting coefficients to arrive at a converged solution, the ACE algorithm guarantees the convergence of the transformations. Once the ACE iteration is converged, a simple regression analysis usually suffices for generating actual analytical functional forms for the transformed variables. Second, the ACE algorithm cannot produce a worse fit than the traditional RSM, because if no transformations are found to be necessary, then the ACE would simply suggest nearly linear transformations for all the variables, indicating at least equal or much greater data fitting than the traditional RSM. While the ACE algorithm provides the several aforementioned advantages, it has not been applied recently to the field of thermal-hydraulic uncertainty analysis [12,13].

The main objective of this paper is to assess the feasibility of the ACE-RSM approach as a complementary method

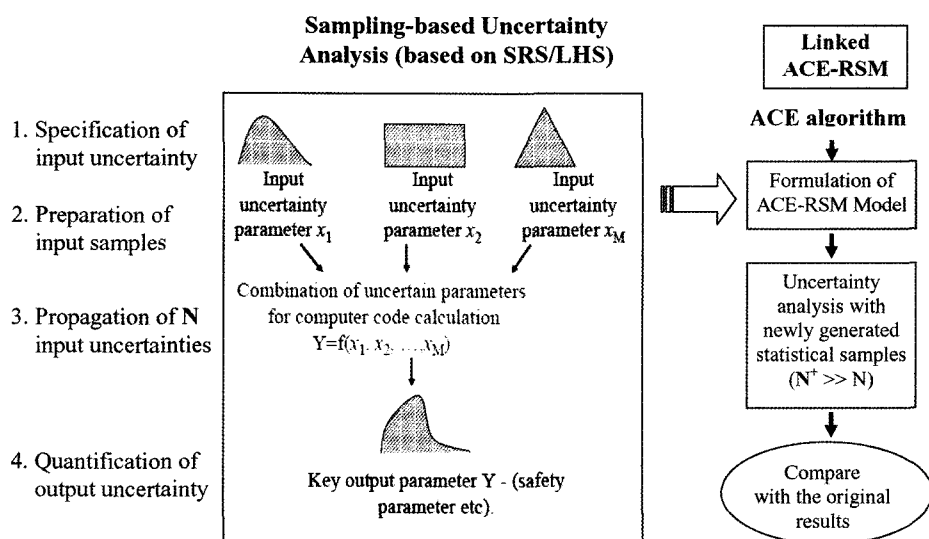


Fig. 1. A Statistical Uncertainty Assessment Process Linked with the ACE-RSM Approach

to the Wilks' formula and the conventional regression-based uncertainty analysis through an example application of it to the OECD BEMUSE Phase III-LOFT L2-5 LBLOCA PCT uncertainty analysis. It should be noted that the term 'uncertainty' mentioned in this paper mainly focuses on the 5%/95% statistical bounds for the calculated PCT values, in particular, the confidence limit in the case that the Wilks' formula is applied, or otherwise, the empirical percentiles. Figure 1 shows the whole process of the ACE-RSM-based LBLOCA PCT uncertainty analysis employed in this paper.

2. BRIEF SUMMARY OF THE ACE ALGORITHM

For a dependent variable and multiple independent variables $(X_i, i=1, \dots, p)$, the objective of the ACE algorithm [8-11] is to find the most optimal transformations $\theta(Y)$ and $\phi_i(X_i)$ that maximize the statistical correlation between $\theta(Y)$ and $\sum_{i=1}^p \phi_i(X_i)$, by treating each value of the transformed variable $\theta(Y)$ as the expectation of several realizations of the sum of transformed independent variables $\sum_{i=1}^p \phi_i(X_i)$. The resulting ACE regression model can be expressed as:

$$\theta(Y) = \alpha_0 + \sum_{i=1}^p \phi_i(X_i) + \varepsilon \quad (1)$$

where θ is a function of the response variable, Y , and ϕ_i are functions of the independent variables, $(X_i, i=1, \dots, p)$, α_0 is the regression constant, and ε is an error term that is not explained by the regression. Thus, the ACE model replaces the problem of estimating a linear function of a p -dimensional variable $X=(X_1, X_2, \dots, X_p)$ by estimating p separate one-dimensional functions, $\phi_i(X_i)$ and $\theta(Y)$ with an iterative method. These transformations are achieved by minimizing the unexplained variance of a linear relationship between the transformed response variable $\theta(Y)$ and the sum of transformed independent variables $\sum_{i=1}^p \phi_i(X_i)$. Eq.(2) indicates the error variance (ε^2) that is not explained by the regression (under the constraint, $E[\theta^2(Y)]=1$):

$$\varepsilon^2(\theta, \phi_1, \phi_2, \dots, \phi_p) = E[\theta(Y) - \sum_{i=1}^p \phi_i(X_i)]^2 \quad (2)$$

The minimization of ε^2 with respect to $\phi_1(X_1), \phi_2(X_2), \dots, \phi_p(X_p)$ and $\theta(Y)$ is carried out through a series of single-function minimization, involving the following conditional expectation and minimization

$$\phi_i(X_i) = E[\theta(Y) - \sum_{j \neq i}^p \phi_j(X_j) | X_i]^2 \quad (3a)$$

$$\theta(Y) = \frac{E[\sum_{i=1}^p \phi_i(X_i) | Y]}{\| E[\sum_{i=1}^p \phi_i(X_i) | Y] \|} \quad (3b)$$

with the square-norm $\|\cdot\|$ introduced such that $E(\theta^2(Y))=1$. That is, the ACE algorithm consists of an iterative use of Eqs.(3) in alternating directions in order to obtain the most optimal transformations of $\phi_i(X_i)$ and $\theta(Y)$, requiring a step-by-step computational job (please see References 8-10 for its formal operation). Then, the transformations $\phi_i(X_i)$ and $\theta(Y)$ after the minimization are estimates of the optimal transformations $\phi_i^*(X_i)$ and $\theta^*(Y)$, leading to the following relationship:

$$\theta^*(Y) = \sum_{i=1}^p \phi_i^*(X_i) + \varepsilon^* \quad (4)$$

Here, ε^* is the error that is not captured by the use of the ACE transformations, which is assumed to have a normal distribution with zero mean. In that case, the minimum regression error, ε^* , and the maximum multiple correlation coefficient, ρ^* , are related by $\varepsilon^{*2} = 1 - \rho^{*2}$. However, the minimization of ε^2 cannot be obtained directly because Eqs.(3a) and (3b) are coupled for each other. Thus, the ACE algorithm necessitates an iterative use of the two smoothing operations of Eqs.(3) in alternating directions in order to obtain transformed independent and dependent variables. When a convergence is attained with an iterative scheme, the data in each transformed variable are usually smooth and they vary slowly. While the ACE-transformation itself does not produce explicit mathematical forms on $Y-\theta^*(Y)$ and $X-\phi_i^*(X_i)$, the ACE transformed data plots generated as a result of the above iteration provides an essential information to estimate those mathematical forms. This means that as a further step, it is necessary to derive explicit mathematical forms for the transformed variables with the data plots. Once the explicit functional forms for the transformations are estimated with the ACE-transformed data plots, we can perform a standard regression analysis for these transformations to obtain $\theta^*(Y) = \alpha_0 + \sum_{i=1}^p \alpha_i \phi_i^*(X_i)$ with the corresponding regression coefficients α_i . Then, the final functional form of Y versus X_1, \dots, X_p , i.e., ACE-RSM model, is given in the form of $Y = \theta^{-1}[\alpha_0 + \sum_{i=1}^p \alpha_i \phi_i^*(X_i)]$. Here, $\theta^{-1}[\cdot]$ indicates an inverse function of $\theta^*(Y)$.

On the other hand, it should also be noted that the error in Eq.(2) could vanish with a judicious choice of $\theta(y)$, if $\theta(y_j)$ equals $\phi_i(x_{i,j})$ for every point in a set of N data $\{(x_{i,j}, y_j), j=1, 2, \dots, N\}$. In practice, however, this idealized situation does not occur because the data contain a greater or lesser randomness and so do $\theta(y_j)$ and $\phi_i(x_{i,j})$. As in conventional regression analysis, thus, $\theta(y_j)$ in the ACE algorithm is also estimated to the expectation of several realizations of $\phi_i(x_i)$ for the j 'th point, rather than a single unique realization $\phi_i(x_{i,j})$. In most regression problems, there is usually only one value y_j , and hence one value $\phi_i(x_{i,j})$, for the j 'th data point, and the conditional expectation $\theta(y_j)$ has to be evaluated with the neighboring values $\{\phi_i(x_{i,k}), k=j-M, \dots, j+M\}$, for some M , which is treated as realizations of $\phi_i(x_{i,j})$, for the j 'th data point. Thus, data

smoothing operation over the dependent and independent variables plays a primary role in the ACE algorithm [8-11].

3. MARS SIMULATION FOR THE PCT UNCERTAINTY ANALYSIS

In order to assess the PCT uncertainty for a LBLOCA blowdown phase, the sampling-based method [14] was applied to the MARS code (version 2.3) [2,15]. As the first step for the uncertainty analysis, 14 uncertainty parameters were chosen based on the result of the Phenomena Identification and Ranking Table (PIRT) on LBLOCA. As given in Table 1, the chosen parameters cover (a) physical models employed in the code and (b) initial and boundary conditions for the LBLOCA simulation. The state of knowledge about all uncertain parameters has been described by the ranges and subjective probability distribution (see also Table 1). Second, the random variance of each uncertain parameter was determined by a crude Monte Carlo sampling method, which is based on the combined probability distribution of the uncertain parameters. For uniform distribution, the minimum and maximum values are the boundaries of sampling. For normal distribution, the sampling boundaries were truncated at its mean value $\pm 2\sigma$. Any dependency between parameters was not considered in the sampling process. Then, the MARS code calculations were performed with sampled sets of parameters. In order to limit the number of samples

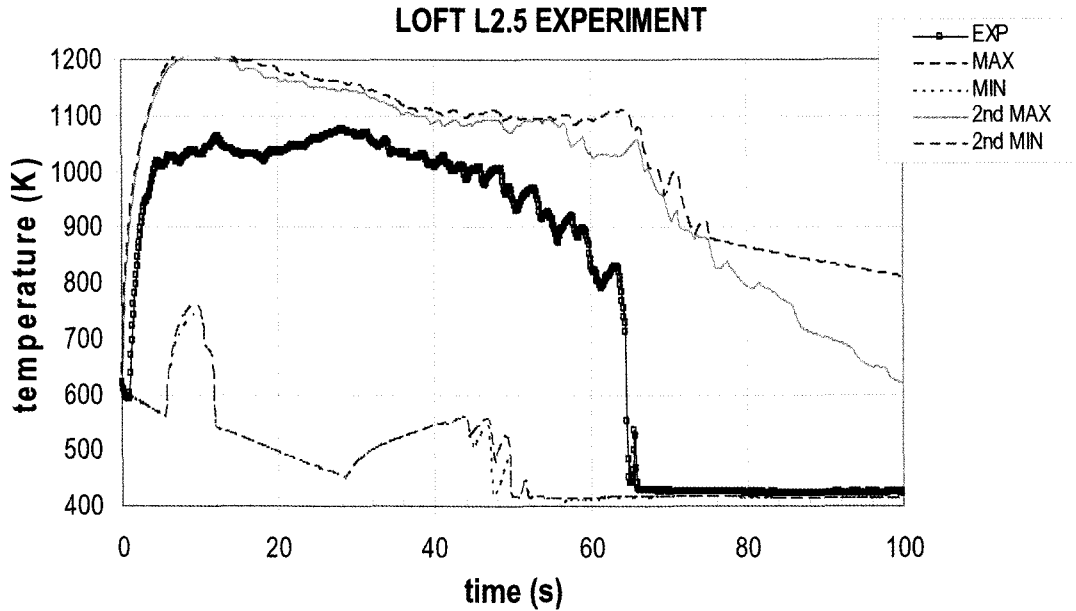
for the MARS code calculations to a reasonable level, a set of 93 samples were required to apply Wilks' formula unilaterally at the 2nd order (for 95% tolerance limit and 95% confidence interval of the code results), but 100 samples were sampled to get sufficient calculation numbers that can cover the unexpected run errors. Finally, multiple input decks (each input deck has 100 statistical inputs) were implemented to identify the effect of different sets of random samples on the PCT value, consequently resulting in several thousands of PCT values for its statistical analysis. The short transient time (i.e., 100 seconds in case of LOFT L2-5) was enabled to perform the multiple sets of these 100 cases.

For each input deck (composed of 100 statistical inputs), the ratio of the failed runs (i.e., crash of the simulation code) was 7% (i.e., 7 out of 100 runs were failed). The failure was random and its reasons were not known (maybe due to a combination of physically unreasonable random values). The failed calculations were discarded manually. The remaining outputs are the final result of the propagation of input uncertainties through the specified number of code calculations. According to a recommendation from the BEMUSE Phases III, in order for the Wilks' formula to be properly used, all the code runs must be successful or corrected in the case of failure. For a small number of code failures, the BEMUSE Phases III also recommends that a conservative approach be used by performing more code runs, which is similar to the present approach. For a typical input deck with 93 samples,

Table 1. MARS Input Parameters Employed for PCT Uncertainty Analysis

Component	Associated Phenomena	Code input x_i	Independent parameters and their ranges: $\pm 2\sigma$ (or min/max)		PDFs
Core	Reflood heat transfer	1	D-B Liquid HT multiplier	$\pm 20\%$	Normal
		2	Chen's Nucleate boiling HT multiplier	$\pm 23.2\%$	Normal
		3	AECL lookup CHF multiplier (Bowdown CHF)	$\pm 74\%$	Normal
		4	Transition boiling multiplier	$\pm 32\%$	Normal
		5	Film boiling HT multiplier	$\pm 36\%$	Normal
		6	D-B Vapor HT multiplier	$\pm 20\%$	Normal
Fuel rod	Stored Energy	7	Peaking factor(Fq)	$\pm 14.96\%$	Normal
		8	Cold gap size	$\pm 20.98 \mu\text{m}$	Uniform
		9	Gap conductance multiplier	$\pm 80\%$	Uniform
		10	Fuel conductivity	$\pm 10\%$	Normal
		Decay Heat	11	Decay heat	$\pm 6.6\%$
Break	Critical flow	12	Break area ratio	0.7~1.15	Uniform
Pump	Two-phase performance	13	Pump two-phase performance	0.0~1.0	Uniform
Down comer	ECC Bypass	14	Downcomer lateral loss coeff.	0.0~1.0	Uniform

Note HT: Heat Transfer; D-B: Dittus-Boelter correlation



[Note] 5%/95% bounding values: 2nd MIN/2nd MAX (based on the 2nd order of Wilks' formula (i.e., 93 samples))

Fig. 2. The 5%/95% PCT Bounding Values (MARS Calculations) Compared with the LOFT L2-5 Experimental Data

Figure 2 shows the 5%/95% PCT bounding of the MARS calculations compared with the LOFT experimental results.

4. FORMULATION OF ACE-RSM MODELS AND RESULT ANALYSIS

Based on the MARS input x_i ($p=14$) and output values for the PCT (y) uncertainty analysis, the corresponding ACE-RSM models can be formulated through the following procedures:

(Step-1): Derive two types of the ACE-transformed functional forms for independent and dependent variables with the prepared N sample input and output values: $x_i \sim \hat{\phi}_i(x_i)$ and $y \sim \hat{\theta}(y)$

(Step-2): Perform a (piece-wise) linear regression between the transformed variables, $\hat{\phi}_i(x_i)$ and $\hat{\theta}(y)$:

$$\hat{\theta}^*(y) = \alpha_0 + \sum_{i=1}^{p=14} \alpha_i \hat{\phi}_i^*(x_i). \quad (5)$$

Here, it should be noted that when a convergence is well attained with an iterative scheme, a regression of the transformed dependent variable on all the transformed covariates results in all of the parameter coefficients of the independent variables ($\alpha_i, i=1,2,\dots,p$) being positive and close to 1 and the regression constant $\alpha_0 \approx 0$.

(Step-3): Derive the final functional form between the original input and output variables, x_i and y

Table 2. Wilks' Statistical Samples Based Confidence and Tolerance Limit ($\alpha = \beta = 95\%$)

Wilks' formula	# of samples (N)	For N ordered output values obtained N random samples, $y_1 < y_2, \dots, < y_N$	
		Rank corresponding to 95% value	Rank corresponding to 5% value
1 st order	59	y_N	y_1
2 nd order	93	y_{N-1}	y_2
3 rd order	124	y_{N-2}	y_3
4 th order	153	y_{N-3}	y_4

Note: Confidence interval bounds a population parameter (e.g., mean) with some confidence, while a tolerance interval bounds percentile range that represents a specified proportion of the population. In simpler terms, the confidence interval characterizes the behavior of the average while the tolerance interval characterizes the behavior of a range of individuals.

$$\hat{y} = \hat{\theta}^{*-1} \left\{ \alpha_0 + \sum_{i=1}^{p=14} \alpha_i \hat{\phi}_i^*(x_i) \right\}. \quad (6)$$

Using the foregoing process, 4 typical ACE-RSM models on the PCT were formulated based on a limited number of sample runs: (1) *one from N=93: ACE-RSM-Model-1*, (2) *two from N=124: ACE-RSM-Model-2(a) (Random Set-1) and ACE-RSM-Model-2(b) (Random Set-*

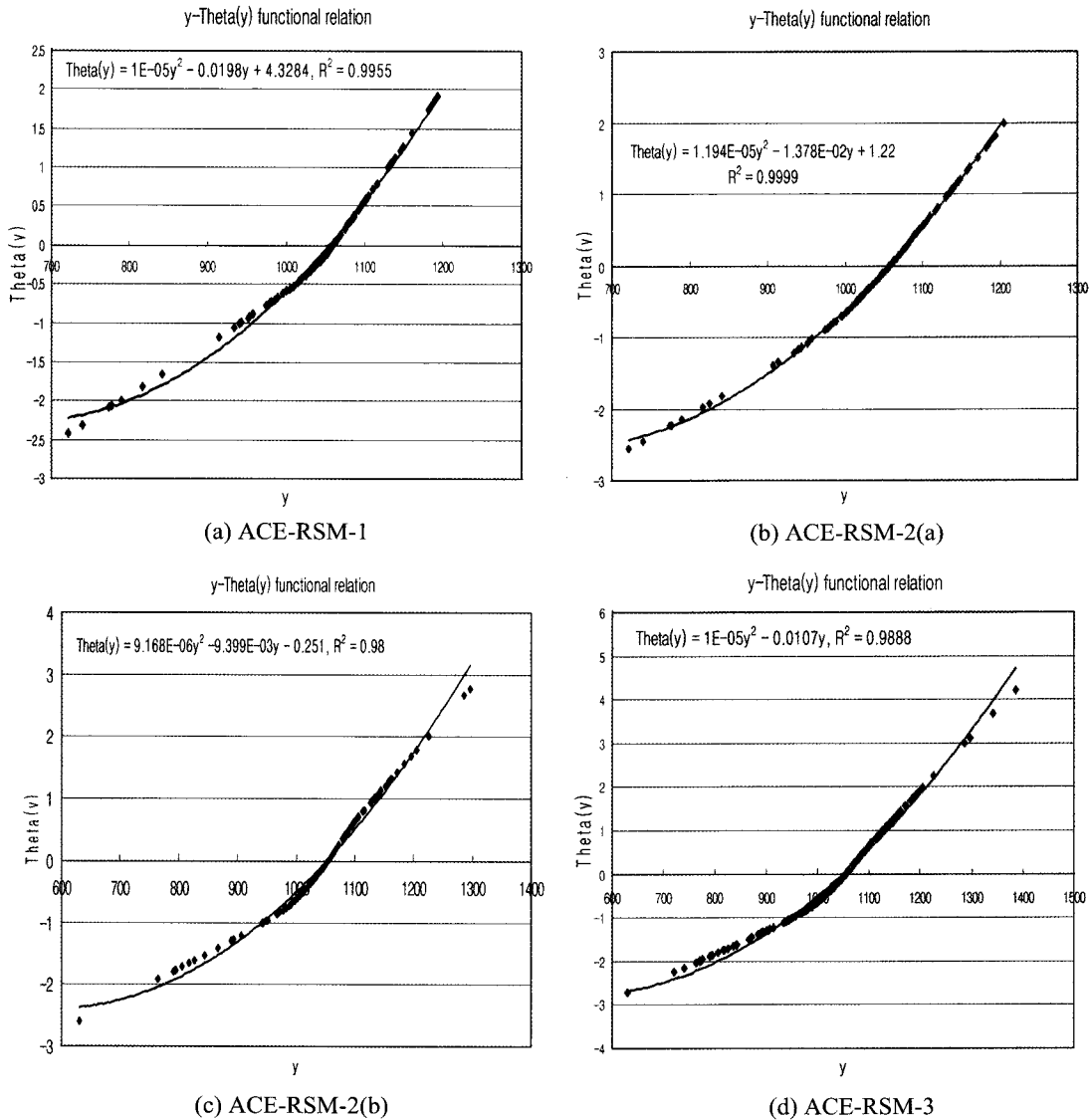


Fig. 3. Functional Forms of the ACE-Transformed PCT Variable $\hat{\theta}(y)$

2), and (3) one from $N=300$: ACE-RSM-Model-3. As shown in Table 2, the number of selected samples, $N=93$ and $N=124$, correspond to the 2nd and 3rd orders of Wilks' formula (subject to a tolerance limit of 95% and a confidence level of 95% for the code results), respectively, and $N=300$ corresponds to twice the number of samples ($N=153$) at the 4th order of Wilks' formula.

As a result of the Step-1, Figure 3 shows the functional relations between x_i and $\hat{\phi}_i(x_i)$ suggested by the ACE plots: the piece-wise linear (or polynomial) format of x_i in the case of $\hat{\phi}_i(x_i)$ and the piece-wise linear (or quadratic) format of y in case of $\hat{\theta}(y)$. As a result of the Step-2, Table 3 and Figure 4 show the ACE-transformed regression coefficients ($\alpha_i, i=1,2,\dots,p$) between $\hat{\phi}_i(x_i)$ and $\hat{\theta}(y)$ and the relative differences between a sum of the transformed

inputs $\sum_{i=1}^{p=14} \phi_i(x_i)$ and the transformed PCT variables $\hat{\theta}(y)$, respectively. As shown in Figure 4, the plots of transformed variables $\hat{\theta}(y)$ and $\hat{\phi}_i^*(x_i)$ that were subjected to a higher R^2 , explain a good regression between $\hat{\theta}(y)$ and $\hat{\phi}_i^*(x_i)$ with the present ACE algorithm. The foregoing results are also justified by the regression coefficients ($\alpha_i, i=1,2,\dots,p$) between $\hat{\phi}_i(x_i)$ and $\hat{\theta}(y)$ in Table 4, approaching 1 for most $\hat{\phi}_i^*(x_i)$. As mentioned before, $\alpha_i=1.0$ means a perfect correlation (or regression) between $\hat{\phi}_i(x_i)$ and $\hat{\theta}(y)$. This is especially true when several kind of linear regressions (including partial correlation and standardized regression [16,17]) were applied, as shown in Table 4, where the resultant R^2 values were much lower than those of the corresponding ACE-RSM cases, thereby explaining a much worse performance for the regression.

Table 3. Regression Coefficients between the ACE-Transformed Inputs and Output

$\hat{\phi}(x_i)$	Regression coeff. (α_i) and the square of the multiple correlation coefficient			
	ACE-RSM-1	ACE-RSM-2(a)	ACE-RSM-2(b)	ACE-RSM-3
$i=0$	$\alpha_0 = -1.73E-07$	$\alpha_0 = -6.93E-07$	$\alpha_0 = -6.93E-07$	$\alpha_0 = -2.03E-06$
$i=1$	$\alpha_1 = 1.08$	$\alpha_1 = 1.17$	$\alpha_1 = 1.67$	$\alpha_1 = 1.05$
$i=2$	$\alpha_2 = 1.09$	$\alpha_2 = 1.08$	$\alpha_2 = 1.39$	$\alpha_2 = 1.50$
$i=3$	$\alpha_3 = 1.08$	$\alpha_3 = 1.27$	$\alpha_3 = 1.17$	$\alpha_3 = 1.14$
$i=4$	$\alpha_4 = 1.02$	$\alpha_4 = 1.08$	$\alpha_4 = 0.909$	$\alpha_4 = 1.01$
$i=5$	$\alpha_5 = 1.04$	$\alpha_5 = 1.18$	$\alpha_5 = 1.04$	$\alpha_5 = 0.987$
$i=6$	$\alpha_6 = 1.22$	$\alpha_6 = 1.21$	$\alpha_6 = 1.48$	$\alpha_6 = 1.17$
$i=7$	$\alpha_7 = 1.02$	$\alpha_7 = 1.00$	$\alpha_7 = 1.02$	$\alpha_7 = 1.01$
$i=8$	$\alpha_8 = 0.99$	$\alpha_8 = 1.01$	$\alpha_8 = 0.979$	$\alpha_8 = 0.993$
$i=9$	$\alpha_9 = 1.00$	$\alpha_9 = 0.979$	$\alpha_9 = 0.970$	$\alpha_9 = 0.985$
$i=10$	$\alpha_{10} = 1.11$	$\alpha_{10} = 1.06$	$\alpha_{10} = 1.10$	$\alpha_{10} = 1.14$
$i=11$	$\alpha_{11} = 1.16$	$\alpha_{11} = 1.14$	$\alpha_{11} = 0.956$	$\alpha_{11} = 1.15$
$i=12$	$\alpha_{12} = 0.99$	$\alpha_{12} = 1.01$	$\alpha_{12} = 1.00$	$\alpha_{12} = 1.01$
$i=13$	$\alpha_{13} = 1.01$	$\alpha_{13} = 1.01$	$\alpha_{13} = 1.07$	$\alpha_{13} = 1.04$
$i=14$	$\alpha_{14} = 1.16$	$\alpha_{14} = 0.992$	$\alpha_{14} = 1.21$	$\alpha_{14} = 1.10$
R^2	0.98	0.96	0.96	0.93

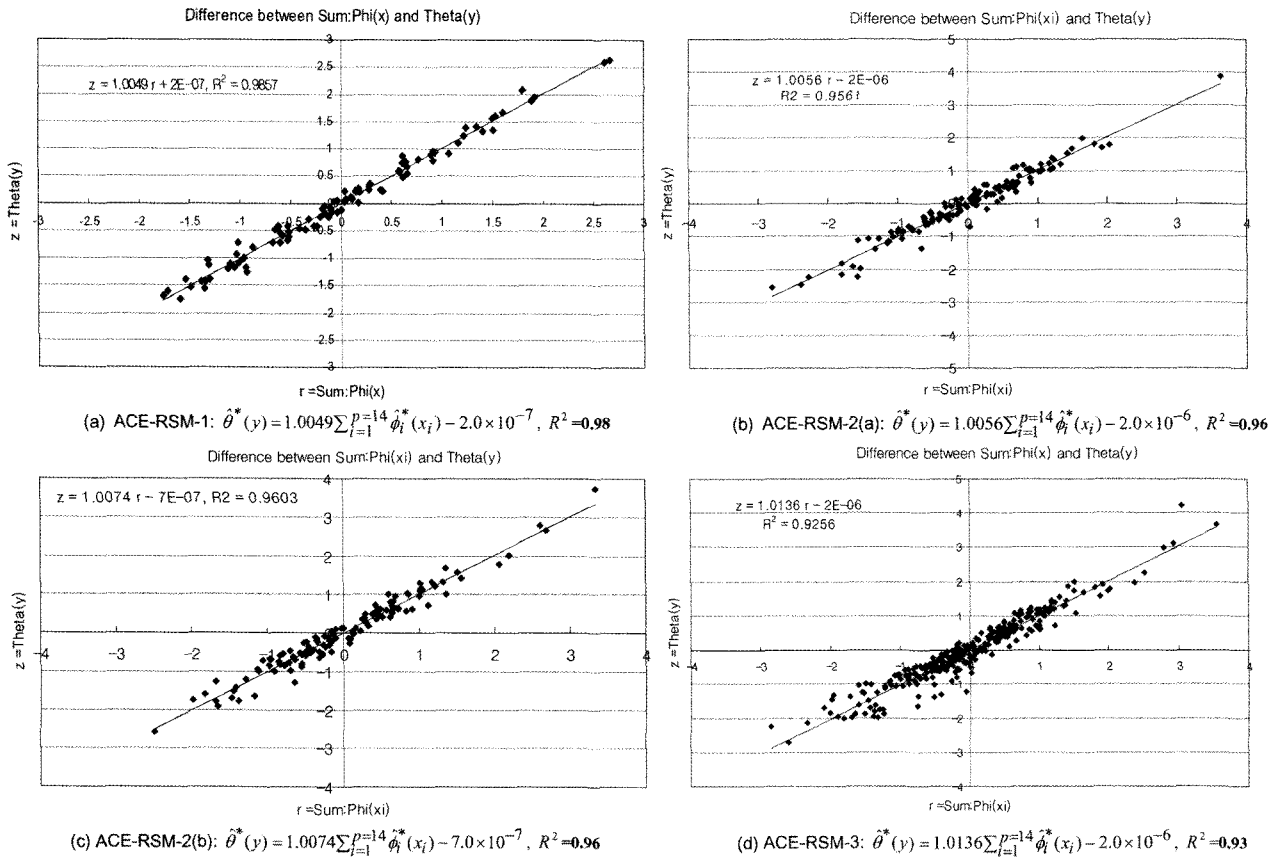


Fig. 4. Relationships between the ACE-Transformed Inputs $\hat{\theta}_i(x_i)$ ($p=14$) and PCT $\hat{\theta}(y)$

Table 4. Linear Regression Coefficients between the Original Inputs and Output

Coefficient of determination	Linear regression between original inputs and output				
	1 st order RSM ⁽¹⁾	Standardized Regression Coeff. (SRC) ⁽²⁾	Partial Correlation Coeff. (PCC)	Standardized Rank Regression Coeff. (SRCC)	Partial rank Correlation Coeff. (PRCC)
R^2 : RSM-1	0.80	0.80	0.80	0.85	0.85
R^2 : RSM-2(a)	0.85	0.85	0.85	0.81	0.81
R^2 : RSM-2(b)	0.84	0.84	0.84	0.82	0.82
R^2 : RSM-3	0.81	0.81	0.81	0.79	0.79

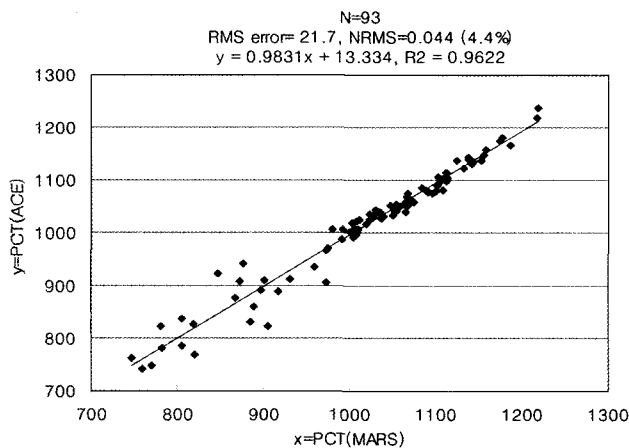
Note

1. Superscripts (1) $\hat{y} = \beta_0 + \sum_{i=1}^{p=14} \beta_i x_i$; (2) $\hat{y}^* = \sum_{i=1}^{p=14} \beta_i^* x_i^*$, $x_i^* = (x_i - \bar{x}_i) / \sigma_i$

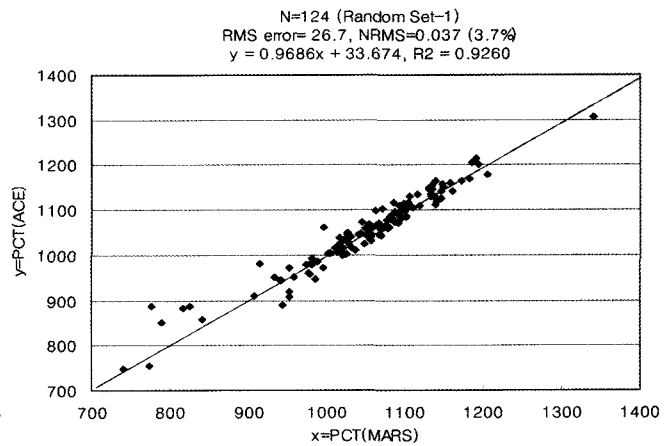
2. SRC: the regression coefficient of standardized data points.

3. PCC: the degree of correlation between two random variables, with the effect of a set of controlling random variables removed

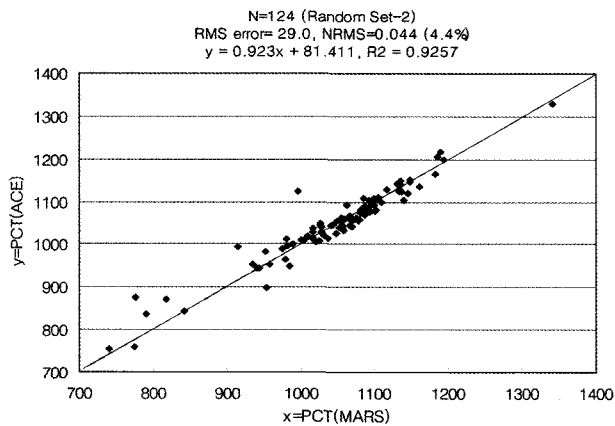
4. SRCC/PRCC: the regression coefficient of ranked data points



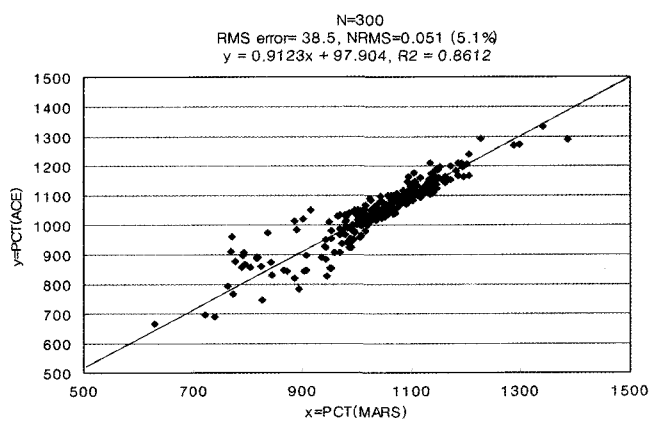
(a) ACE-Model-1: $R^2=0.9622$, Normalized RMS=0.044 (4.4%)



(b) ACE-Model-2(a): $R^2=0.9260$, Normalized RMS=0.037 (3.7%)



(c) ACE-Model-2(b): $R^2=0.9257$, Normalized RMS=0.044 (4.4%)



(d) ACE-RSM-3: $R^2=0.8612$, Normalized RMS=0.051 (5.1%)

Fig. 5. Comparison of the ACE-RSM Results with the Corresponding MARS Results

As a result of the *Step-3*, Figure 5 shows a one-to-one comparison of the ACE-RSM-based PCT for the four ACE-RSM models with the corresponding MARS results based on the corresponding R^2 and Root Mean Square (RMS) (or normalized RMS) error as the performance criteria on the regression.

$$\text{Coefficient of Determination: } R^2 = \frac{\sum_{j=1}^N (\hat{y}_j - \bar{y})^2}{\sum_{j=1}^N (y_j - \bar{y})^2} \quad (7)$$

$$\text{RMS Error: } RMS(\theta_1, \theta_2) = \sqrt{\frac{E[(\theta_1 - \theta_2)^2]}{\sum_{i=1}^N (y_{1,i} - y_{2,i})^2 / N}} \quad (8)$$

$$\begin{aligned} \text{Normalized RMS Error: } NRMS(\theta_1, \theta_2) \\ = RMS(\theta_1, \theta_2) / (y_{ref,max} - y_{ref,min}) \end{aligned} \quad (9)$$

Table 5 shows the 5% and 95% PCT values for the ACE-RSM models and the corresponding MARS values with a different set of samples.

A careful investigation of Figure 5 and Table 5 shows that with the exception of the lower PCT values, (which is subject to highly nonlinear thermal-hydraulic behavior), the ACE-RSM PCT traced relatively well the corresponding MARS results. From the qualitative aspect, the ACE-RSM-Model-3 (based on a larger number of samples) was quickly converged to the original results, compared to the ACE-RSM-Model-2(a) and Model-2(b) that were subjected to higher values of R^2 . This fact indicates that although the

RSM itself is well fitted to the original code results, a smaller number of samples may not trace the whole trend of the original MARS code results that were not explained due to their random variation. In order to cover such a random variation that may not be explained by the limited number of samples, it is better to employ an appropriate number of samples. From the quantitative aspect, the accuracy bound for ACE- Model-3 was in between $\Delta T|MARS-ACE|=5K$ for 95% PCT and 20K for 5% PCT (e.g., see the case of $N=3000$ of Table 5). The above results indicated that the ACE-RSM models could be an appropriate surrogate model to the original MARS code, although they also showed a greater or lesser dependency on the utilized number of samples, as in the conventional RSM.

Finally, a variation of the Cumulative Distribution Function (CDF) according to the increase of sample size was investigated in order to estimate the final form of the corresponding probability distribution. Figure 6 shows that in the case of ACE-RSM-2(a), the PCT approaches the normal PDF as the sample size N increases from 124 to 3000.

5. SUMMARY AND CONCLUSIONS

In this paper, the ACE-RSM approach was applied to assess the blowdown PCT uncertainty for the LOFT L2-5 Experiment as a complementary work to the OECD BEMUSE Phase-III program. Two decision parameters were taken into account in order to assess the performance of the estimated ACE-RSM models: R^2 and RMS/NRMS errors. While the R^2 summarizes the information on the general tendency (i.e., average behaviour) of the estimated regression models, the RMS/NRMS errors cover the impact

Table 5. 5%/95% Bounds on PCT (K) Values Based on the ACE-RSM Models

N	MARS		ACE-RSM-1 ($R^2=0.9622$)		ACE-RSM-2(a) ($R^2=0.9260$)		ACE-RSM-2(b) ($R^2=0.9257$)		ACE-RSM-3 ($R^2=0.8612$)	
	5%	95%	5%	95%	5%	95%	5%	95%	5%	95%
93 ⁽¹⁾	759.2	1217.5	747.0	1218.7	-	-	-	-	-	-
124 ⁽²⁾	790.0	1286.5	-	-	824.3	1200.8	846.1	1212.0	794.0	1204.8
300 ⁽³⁾	817.7	1171.3	-	-	884.8	1189.5	886.9	1178.5	852.6	1182.2
500	805.8	1172.1	-	-	881.7	1182.6	883.0	1176.6	846.7	1173.4
1000	810.7	1173.5	-	-	864.8	1181.4	867.7	1177.4	835.0	1173.7
3000	815.5	1174.4	-	-	861.2	1179.0	867.8	1177.7	834.4	1176.0
6000	N/A	N/A	-	-	859.3	1179.3	867.5	1178.6	834.8	1177.1

Note superscript

- (1) Based on the 2nd order of Wilks' formula ($N=93$), i.e., 5% bound = $y(2)$ & 95% bound = $y(92)$.
- (2) Based on the 3rd order of Wilks' formula ($N=124$), i.e., 5% bound = $y(3)$ & 95% bound = $y(122)$.
- (3) Except for (1) and (2), the remaining ones are based on the 5%/95% empirical percentiles.

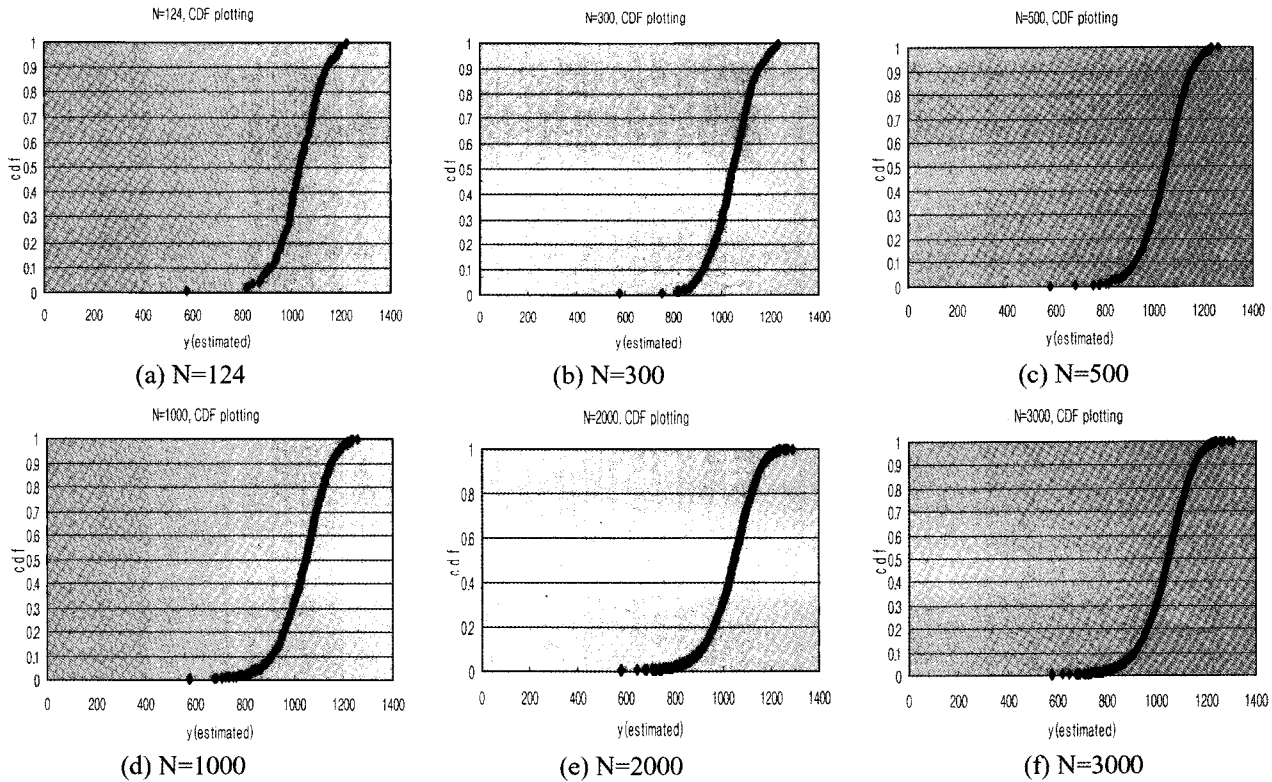


Fig. 6. An Example of CDF Variation According to the Increase of Samples (ACE-RSM-2(a) Case)

of local variations.

A comparison with the original MARS code results showed that except for the lower (or 5%) PCT values (which was subject to highly nonlinear thermal-hydraulic behaviour), the formulated ACE-RSM Models relatively traced the original code results well, even with a limited number of code runs. Since the aforementioned nonlinear thermal-hydraulic behavior is highly dependent on the physical domain of a system, rather than a random error, its impact on the final results should be explained by other means in more detail (e.g., physical models employed in the thermal-hydraulic system code). In addition, it should be noted that since the ACE-RSM approach was also a kind of regression technique, its performance might be not good when data points are scattered in highly nonlinear ways. In that case, the best way to obtain a better performance (i.e., curve fittings) is to control data smoothing in various ways.

Based on the foregoing results, the present ACE-RSM approach could be effectively utilized in the field of best estimate and uncertainty analysis, with a balance of accuracy/robustness in predicted uncertainty and computational demands if the data smoothing factor is appropriately controlled. A greater or lesser dependency of the regression performance on the number of samples could be reduced further by employing more efficient

sampling schemes, such as the Latin Hypercube Sampling (LHS) approach [18,19]. The LHS method selects N values randomly from N equally spaced intervals of the cumulative distribution, thereby ensuring good sampling of the distribution tails with a minimum number of samples, and much less dependency on the size of statistical samples than random sampling.

ACKNOWLEDGEMENTS

This work was supported by the Nuclear Research & Development Program of the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korean government (MEST).

NOMENCLATURE

- $E[\cdot]$ Expectation Operator
- $E[\cdot|Y]$ Expectation Conditional on Y
- M Neighboring Data Index (or Value)
- N Number of Data Points (or Samples)
- p Number of Input Variables
- R^2 Coefficient of Determination (Regression)
- $X_i; x_i$ Original Uncertainty Input Variables
- $Y; y$ Original Input Variables

Greek Symbols

- α Confidence limit of the Wilks' formula

α_i	Regression Coefficient for the <i>i</i> th ACE-transformed Variable
β	Tolerance limit of the Wilks' formula
β_i	Regression Coefficients for the <i>i</i> th Original Variable
ε	Regression Error Term
$\phi(X)$	ACE-transformed Input Variables (= <i>phi(x)</i> in Figures)
$\theta(Y)$	ACE-transformed Output Variable (= <i>Theta(y)</i> in Figures)
θ_1, θ_2	Two Variables Used in Evaluation of RMS Error
ρ	Multiple Correlation Coefficients
σ	Standard Deviation
$\ \cdot \ $	Square Norm (Normalization Operator)

Superscripts

*	Variables after Optimal transformation
\wedge	Estimated Form of Transformed Variables
-	Average of Data Points
$^{-1}$	Inverse (Function) Operator

Subscripts

max	Maximum Value
min	Minimum Value
ref	Reference Variable

ABBREVIATION

ACE	Alternating Conditional Expectation
BE	Best Estimate
BEMUSE	Best Estimate Methods & Uncertainty and Sensitivity Evaluation
CDF	Cumulative Distribution Function
CHF	Critical Heat Flux
CSNI	Committee on the Safety of Nuclear Installations
HT	Heat Transfer
LBLOCA	Large Break Loss of Coolant Accident
LHS	Latin Hypercube Sampling
LOCA	Loss of Coolant Accident
LOFT	Loss Of Fluid Test
MARS	Multi-Dimensional Analysis of Reactor Safety
NRMS	Normalized Root Mean Square
OECD	Organisation for Economic Cooperation and Development
PCC	Partial Correlation Coefficient
PCT	Peak Cladding Temperature
PDF	Probability Density Function
PIRT	Phenomena Identification and Ranking Table
PRCC	Partial Rank Correlation Coefficient
PWR	Pressurized Water Reactor
RMS	Root Mean Square
RSM	Response Surface Model
SRC	Standardized Regression Coefficient
SRCC	Standardized Rank Regression Coefficient
SRS	Simple Random Sampling

REFERENCES

[1] A. Petruzzi and F. D'Auria, "The BEMUSE Programme:

Best-Estimate Methods Uncertainty and Sensitivity Evaluation - Phase 2," OECD/CSNI Workshop on Evaluation of Uncertainties in Relation to Severe Accidents and Level 2 PSA, Aix-en-Provence, France (November 2005).

[2] A. Crécy and P. Bazin, "BEMUSE Phase III Report: Uncertainty and Sensitivity Analysis of the LOFT L2-5 Test," NEA/CSNI/R(2007)4, OECD, October 11 (2007).

[3] S.S. Wilks, "Determination of Sample Sizes for Setting Tolerance Limits," *The Annals of Mathematical Statistics*, 12, 91-96 (1941).

[4] W.T. Nutt and G.B. Wallis, "Evaluation of Nuclear Safety from the Outputs of Computer Codes," *Reliability Engineering and System Safety*, 83, 57-77 (2004).

[5] J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn, "Design and Analysis of Computer Experiments," *Statistical Science*, 4(4), 409-435 (1989).

[6] R. H. Myers and D. C. Montgomery, *Response Surface Methodology*, John Wiley & Sons, New York (1995).

[7] Y. A. Hassan and C. Fu, "RELAP50MOD3.2 Analysis of a VVER-1000 Reactor with UO₂ Fuel and Mixed-Oxide Fuel," *Nuclear Technology*, 148, 325-334 (2004).

[8] W.S. Cleveland, Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of American Statistical Association*, 74, 828-836 (1979).

[9] L. Breiman and J.H. Friedman, "Estimating Optimal Transformations for Multiple Regression and Correlation," *Journal of American Statistical Association*, 80, 580-598 (1985).

[10] H.G. Kim and J.C. Lee, "Development of a Generalized Critical Heat Flux Correlation through the Alternating Conditional Expectation Algorithm," *Nuclear Science and Engineering*, 127, 300-316 (1997).

[11] D. Wang and M. Murphy, "Identifying Nonlinear Relationships in Regression using the ACE Algorithm," *Journal of Applied Statistics*, 32(3), 243-258 (2005).

[12] H. Glaeser, "Uncertainty Evaluation of Thermal-hydraulic Code Results," International Meeting on "Best-Estimate" Methods in Nuclear Installation Safety Analysis (BE-2000), Washington, DC, November (2000).

[13] IAEA, "Best Estimate Safety Analysis for Nuclear Power Plants: Uncertainty Evaluation, IAEA Safety Reports Series No.52, Vienna, Austria (1984).

[14] J.C. Helton and F.J. Davis, "Illustration of Sampling-Based Methods for Uncertainty and Sensitivity Analysis, *Risk Analysis*, 22(3), 591-622 (2002).

[15] B.D. Chung, "Uncertainty Quantification of LOFT L2-5 Experiment," OECD/NEA BEMUSE Phase III Activity 3rd Meeting, Grenoble, France, October 26-28 (2005).

[16] R.L. Iman, J.C. Helton, and J.E. Campbell, "An Approach to Sensitivity Analysis of Computer Models: Part I- Introduction, Input variable selection and Preliminary variable assessment," *Journal of Quality Technology*, 13, 174-183 (1981).

[17] R.L. Iman, J.C. Helton, and J.E. Campbell, "An Approach to Sensitivity Analysis of Computer Models: Part II- Ranking of input variables, Response surface validation, Distribution effect and technique synopsis," *Journal of Quality Technology*, 13, 232-240 (1981).

[18] R.L. Iman and M.J. Shortencarier, "A Fortran 77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use with Computer Models,

NUREG/CR-3624, Technical Report SAND83-2365, Sandia National Laboratories, Albuquerque, NM, USA (1984).
[19] M.D. McKay, R.J. Bechman, and W.J. Conover, "A

Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, **21**(2), 239-245 (1979).