Transactions of the Korean Nuclear Society Autumn Meeting
Changwon, Korea, October 20-21, 2022

Preliminary Study on Nuclear System Analysis Code Improving Methodology using Machine Learning Technique and Non-Parametric Statistics Theory

ChoHwan Oh, Jeong Ik lee*
Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology (KAIST)
**Corresponding author: jeongiklee@kaist.ac.kr*

1. Introduction

Nuclear reactor is an enormous equipment operating at high temperature and pressure, which is very difficult to conduct accident study at full-scale. A reactor accident experiment is carried out in a scale-down facility, and the full-scale accident results are predicted with a reactor safety analysis code analysis. In the process of the nuclear power plant licensing, the safety of the reactor is comprehensively judged through the verified safety analysis code. Most safety analysis codes contain governing equations and constitutive equations. The constitutive equations have a major impact on the code's accuracy. In the constitutive equations, different equations are used depending on the heat transfer or flow regime. Most of the equations are correlations generated from regression of the separate effect test (SET) data, which is tailored to predict a specific thermal-hydraulic phenomenon well. Moreover, an arbitrarily determined value by a programmer is included from the engineering judgement to maintain numerical stability in the code [1].

Recently, integral effect tests (IETs) are conducted in the world. It represents the primary and the secondary coolant systems. The safety analysis codes are used for the prediction of IETs. ATLAS-DSP is one of the integral effect tests conducted by KAERI and KINS [2]. In this process, it showed that the code accuracy is relatively unsatisfactory for IET compared to the SET test. Many times, constitutive equations were modified to have better IET results. For example, an alternative interfacial drag model was recommended to have better accuracy for prediction in the downcomer during DSP-01 campaign [3].

However, to modify constitutive equation, it is not easy to directly use IET data. IET has many instruments and the thermal-hydraulic range and time scale are large. Furthermore, spatial and temporal dependencies in the calculation results make evaluation of the overall effect on the code result difficult, and thus selecting a constitutive equation and determining the direction for the modification are challenging. There are some studies on the sensitivity and uncertainty analyses of the constitutive equations for the entire range of each heat transfer and flow regimes. However, the constitutive equations are not biased in the same direction in reality. Thus, SET data sets used to generate the constitutive equations does not cover the entire regime. Some parts of the equations have high accuracy with some part has low accuracy.

Therefore, in this study, constitutive equation analysis and weakness identification through the constitutive equation regime subdivision is going to be conducted. The progress of the final study is as follows: regime subdivision and constitutive equation analysis. First, the constitutive equation data for regime subdivision is generated. Data is generated in wide possible range to represent each regime. Then, using the generated data, regime subdivision is performed. In this process, self-organizing map (SOM) clustering, a machine learning technique, is used. For the preliminary study, data generation and data clustering are covered in this paper. In the future, the constitutive equations analysis is conducted by applying the multiplier coefficient to the subdivided regime. The Latin hypercube sampling method is used to generate the multiplier coefficient sets, and constitutive equations are analyzed based on the non-parametric statistics. In previous studies [4], similar study was conducted at the single IET experiment (DSP-05) as target experiment. In this study, multiple IET experiments is going to be used as the target experiment for the constitutive equation analysis. Also, there is a modification of the constitutive equation data used for data clustering It is necessary to regenerate the trained self-organizing map and clusters which is different result from the previous study.

2. Data Generation

Data is generated using the MARS-KS constitutive equation. The data generation method and thermal-hydraulic range is same as in the previous study [4]. The constitutive equation is the function of thermal-hydraulic parameters, geometry and material property. The input parameters (i.e. thermal-hydraulic parameters, geometry and material information) is randomly selected in the given range, and calculate the constitutive equations by applying the data as an input to the MARS-KS constitutive equations. The range is set to include most of the design basis accident of the APR1400. Table I shows the range of the input parameters.

Table I: Problem Description

Input parameters	Range	Unit
Pressure	0.09 – 19	MPa
Fluid temperature	25 – (Tsat+50)	℃
Wall temperature	25 –1184	℃
Void fraction	0 – 1	
Mass flux	108 – 5400	kg m^{-2} s^{-1}
Slip ratio	1 – 3	
Hydraulic diameter	8E-4 – 12	m

Volume length	0.01 – 550	m
Angle	0 or 90	° (degree)
Roughness	0 – 2.0E-4	m

In the input parameter sampling process, most of the parameters are selected with uniform random distribution. However, some parameters are sampled with different method to reflect the two-phase characteristics. In the two-phase flow, most of the fluid temperature is near the saturation temperature. To reflect this phenomenon, the fluid temperature is generated in three different ways: log random distribution, uniform random distribution, and single value. The nucleate regime is generated in the case that the wall temperature is slightly higher than the saturation temperature. Therefore, log random distribution is used when the wall temperature is higher than saturation temperature. In the void fraction generation, uniform random distribution and single value are used to generate the zero and unity value. In the case of geometry and material information, data is generated using a single value and uniform random distribution because only specific values are used in the accident simulation. Fig. 1 shows the distribution used to generate the data.

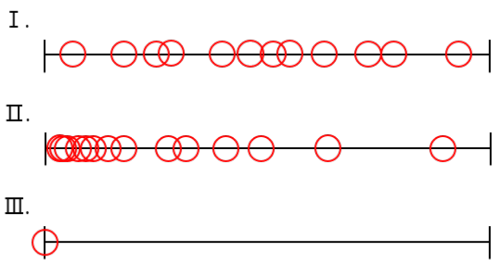

Fig. 1. Distribution used to generate the data (I: uniform random distribution, II: log random distribution, III: single value)

As in the previous study [4], the same number of data for each regime is generated. In the previous study, there were nine regimes in the flow regime. Stratified regime was not sub-divided to the vertical stratified and horizontal stratified regime. Since vertical stratified regime uses different constitutive equation with horizontal stratified regime, stratified regime is separated with vertical and horizontal stratified regime in this study. That is, total ten regimes are in the flow regime. When data is concentrated in a specific regime, it may cause a bias according to the number of data during the clustering process. Therefore, the same number of data for each regime is generated. At least ten-thousand for each regime were generated to cover as wide range as possible. An example of the generated data is shown in the Fig. 2.

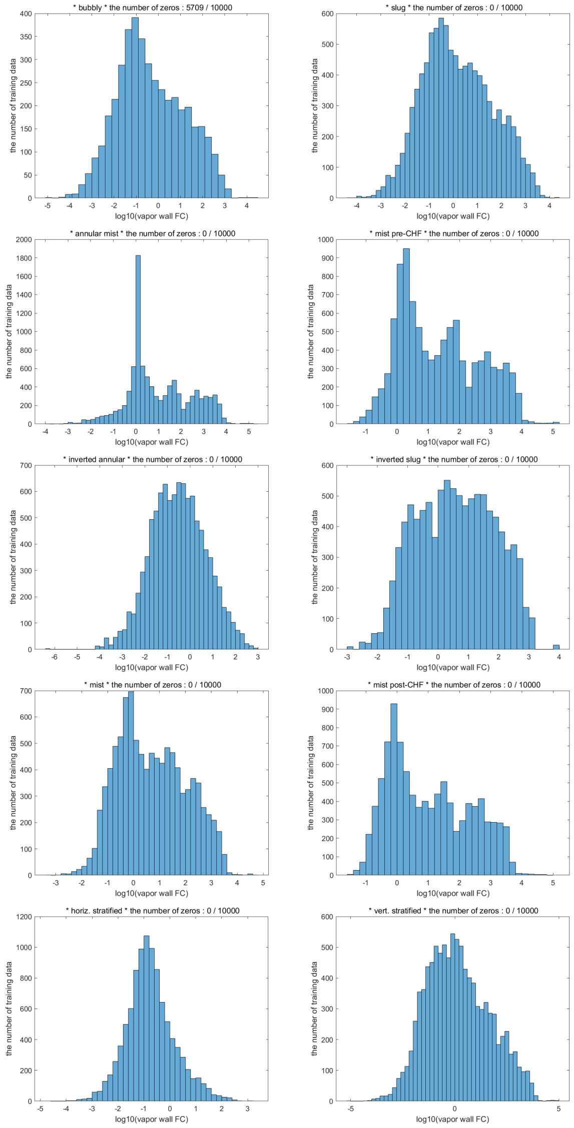

Fig. 2. Results of the constitutive equation generation: vapor wall friction data

3. Data Clustering

Data clustering is performed to subdivide the existing regime map in the MARS-KS code. The parameters used for clustering is the constitutive equations for each regime and the heat/flow regime. Since the constitutive equation values have a very wide range, the logarithm value of the constitutive equation is normalized in [0,1]. The data applied to clustering is in Fig. 3.

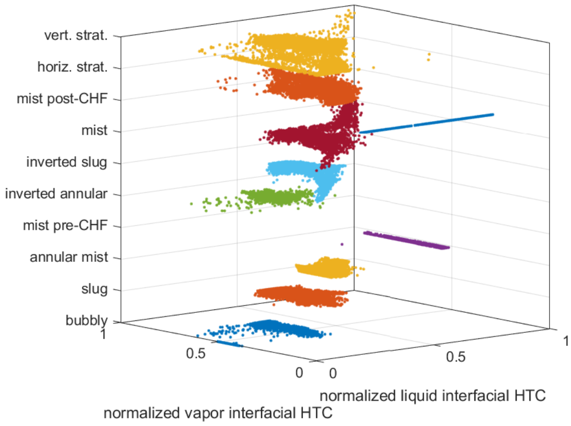

Fig. 3. Wall friction data applied to the SOM clustering.

SOM clustering methodology is used for regime map subdivision [5]. SOM clustering is calculated in following process: SOM mapping and k-means method. First, self-organizing map is trained to replicate the training data. After training, SOM node is clustered with the clustering method. In this study, k-means methodology is used for SOM node clustering which is widely used in various study [6]. It has advantage on the computational cost in the large data sets compared to the other clustering methods [5]. SOM mapping algorithm is in Fig. 4, and the hyperparameters for the SOM mapping is in Table II.

Table II: Problem Description

Map size	30×30
Initial topology	Hexagonal layer
The number of iterations	10,000
Learning resistant	$1 - t/t_{\text{end}}$

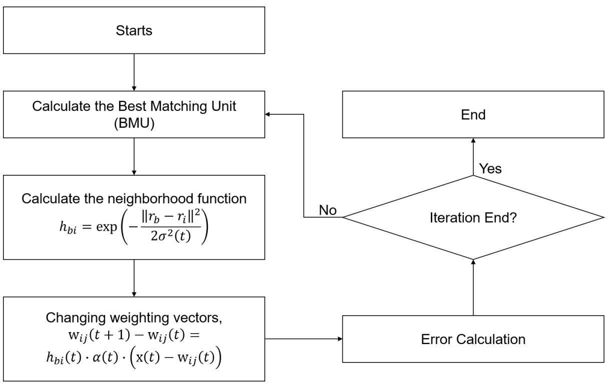

Fig. 4. SOM mapping algorithm

After the SOM mapping, k-means methodology is applied. Initial clusters are randomly determined in the k-means, and iteration is performed until the cluster results does not change. In this methodology, the number of clusters should be predetermined. Therefore, k-means methodology is conducted for each clustering number, and optimal clustering number is calculated with the clustering indices. The average silhouette coefficient [7], Dunn index [8], and Davies-Bouldin index [9] are used for the clustering indices. The clustering results is optimal when the average silhouette coefficient and Dunn index is higher. In contrast, lower value of the Davis-Bouldin index is high quality clusters. The equation and logic of each index is detailed in the references.

As the main purpose of this process is to subdivide the regime, the minimum cluster number is determined to divide the existing regime at least two sub-regimes. Wall heat transfer and wall friction is clustered, and the regime subdivision of interfacial heat transfer and friction is currently in progress. Fig. 5 shows the clustering results of the wall friction.

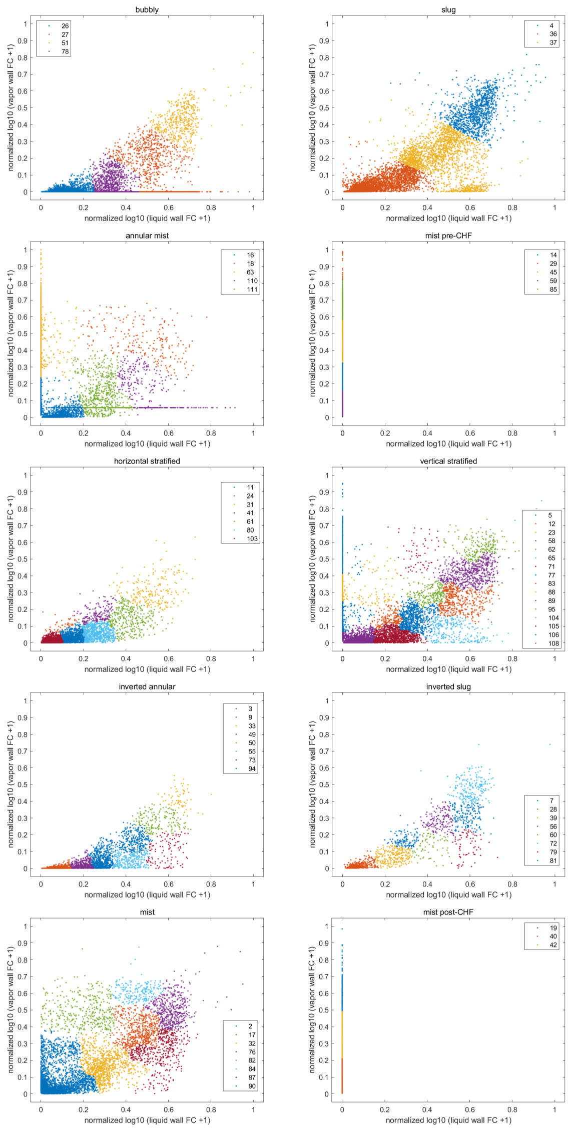

Fig. 5. SOM mapping results: wall friction

4. Summary and Further Works

In this study, a methodology for improving nuclear safety analysis code performance is developed using machine learning technique and non-parametric statistics. The constitutive equation data which can represent each regime is generated, and the regime is subdivided using the SOM clustering method. In the data generation process, the thermal-hydraulic, geometry and material information range is considered to include the design basis accident of APR 1400 as possible, which reflect the two-phase characteristic. Also, clustering indices are used to find the optimal clustering number.

In the future, SOM clustering will be applied to the interfacial heat transfer and interfacial friction. For the constitutive equation analysis, multiplier coefficient sets which applied to each sub-regime are selected using the Latin hypercube sampling. The number of sets is 59 which is based on the Wilks' formula. Sensitivity analysis and uncertainty analysis will be conducted on constitutive equations. Target experiments are DSP-05 and DSP-06. The results will be compared with the single IET results which is performed in the previous study [4].

ACKNOWLEDGEMENT

This work was supported by the Nuclear Safety Research Program through the Korea Foundation Of Nuclear Safety(KoFONS) using the financial resource granted by the Nuclear Safety and Security Commission(NSSC) of the Republic of Korea. (No. 1903002)

REFERENCES

[1] KAERI, MARS CODE MANUAL Vol. 5 Models and Correlations, and Solution Methods, 2009.
[2] Lee, J. B. Description Report of ATLAS Facility and Instrumentation (Second Revision). KAERI/TR-7218/2018, Korea Atomic Energy Research Institute, 2018.
[3] Y.S. Kim et al., First ATLAS Domestic Standard Problem (DSP-01) for the Code Assessment, Nuclear Engineering and Technology, Vol. 43, Issue 1, pp. 25-44, 2011.
[4] ChoHwan Oh et al., Improvement Methodology of Constitutive Equations in Safety Analysis Code using Integral Effect Test Data, Transactions of the Korean Nuclear Society Spring Meeting, Jeju, Korea, May 19-20, 2022.
[5] J. Vesanto and E. Alhoniemi, Clustering of the Self-Organizing Map, IEEE Transaction on Neural Networks, Vol. 11, Issue 3, pp. 586-600, 2000.
[6] J. MacQueen, Some Methods for Classification and Analysis of Multivariate Observations, in: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, pp. 281-297, 1967.
[7] Peter J. Rousseeuw, Silhouette: a graphical aid to the interpretation and avlidation of cluster analysis, Journal of Computational and Applied Mathematics, Vol. 20, pp. 53-65, 1987.
[8] J. C. Dunn, A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, Journal of Cybernetics, Vol. 3, Issue 3, pp. 32-57, 1973.
[9] David L. Davies and Donald W. Bouldin, A Cluster Separation Measure, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-1, No. 2, pp. 224-227, 1979.