

Combinatorial Reduction of Multi-Unit Source Terms Using Clustering Technique

Seunghyeon Hwang, Jaehyun Cho*

Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul, 06974 Republic of Korea

*Corresponding author: jcho@cau.ac.kr

***Keywords :** multi-unit PSA, source term combination, k-medoids clustering, MELCOR, time-dependent release characteristics

1. Introduction

Most nuclear power plant sites worldwide consist of multiple reactor units. Accordingly, site-level risk assessment may need to consider accident scenarios involving more than one reactor unit. In probabilistic safety assessment (PSA), Level 1 analysis identifies core-damage sequences and quantifies their frequencies, while Level 2 analysis evaluates severe accident progression and radioactive release characteristics. The resulting source terms are categorized into source term categories (STCs) and used as inputs for Level 3 offsite consequence analysis. When multiple units experience core damage, combinations of unit-specific STCs need to be considered, and the number of possible combinations can increase substantially with the number of affected units and STCs [1,2].

Several approaches have been proposed to reduce multi-unit STC combinations using logic-tree- or consequence-based grouping criteria [3,4]. Machine-learning-based clustering has also been applied to source term data from single-unit Level 2 PSA to categorize accident scenarios with similar release characteristics [5]. In multi-unit scenarios, source terms may differ in release magnitude, radionuclide composition, and temporal behavior, and releases from individual units may occur at different times or overlap. These characteristics provide a multidimensional basis for grouping multi-unit source term combinations.

In this study, Level 1 and Level 2 PSA results were integrated with source term information obtained from MELCOR analyses, and a methodology based on clustering was applied to reduce the combinatorial complexity of multi-unit source terms. The methodology was developed to identify representative combinations for subsequent offsite consequence analysis.

2. Methodology

The overall procedure of the proposed combinatorial reduction methodology is illustrated in Fig. 1. The workflow consists of multi-unit STC combination generation and aggregation using Level 1 and Level 2 PSA results, source term characterization using MELCOR-derived release histories, and K-medoids clustering followed by representative combination selection.

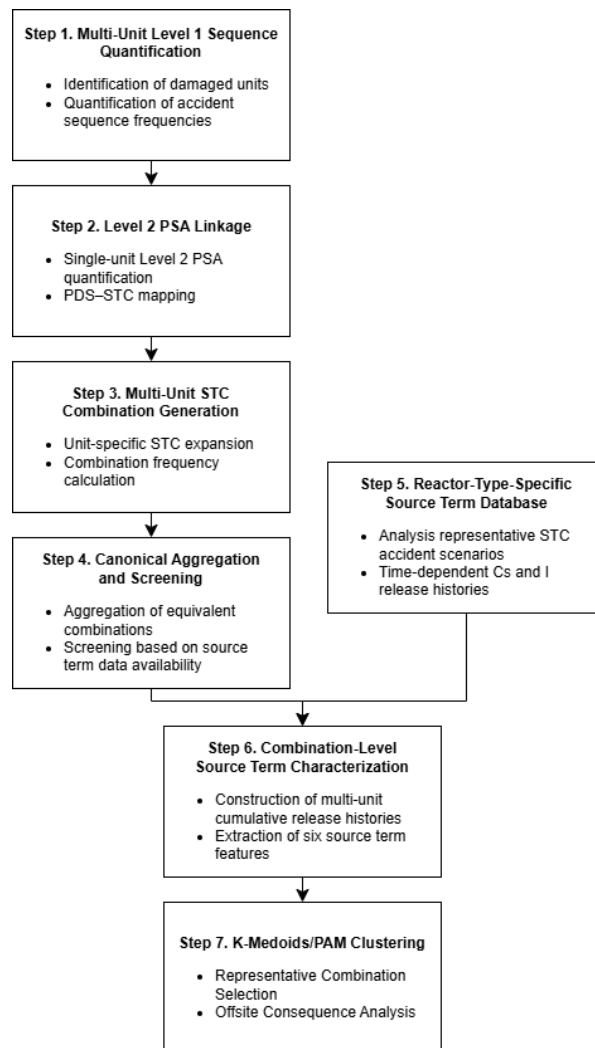


Fig. 1. Overall procedure for combinatorial reduction of multi-unit source terms.

2.1 Generation and Aggregation of Multi-Unit STC Combinations

The methodology was demonstrated using a Hanul nuclear power plant site case-study model based on existing multi-unit Level 1 and Level 2 PSA results. The present case study focused on a multi-unit loss of offsite power (MULOOP) event. The case-study model represents a ten-unit Hanul site configuration consisting of two FRAMATOME, four OPR1000, and four APR1400 units.

Information on damaged units and accident sequence frequencies obtained from the Level 1 PSA was linked with plant damage state (PDS)-STC mapping results from the Level 2 PSA [6]. For each damaged unit, possible STC outcomes and their conditional mapping probabilities were expanded and combined to generate multi-unit STC combinations and their frequency contributions.

Among the MULOOP Level 1 PSA results, 81,941 accident sequences involving core damage in one or more units were retained. The PDS-STC mapping generated approximately 9.46 million raw, unit-specific STC combination records. For source term characterization, units of the same reactor type were treated as exchangeable, and equivalent combinations differing only in the ordering of identical reactor units were aggregated into a single canonical combination, with their frequency contributions summed. After excluding combinations containing STCs for which representative source term data were unavailable, 7,832 canonical STC combinations were retained for clustering.

2.2 Construction of Source Term Features

A reactor-type-specific source term database was constructed using time-dependent radioactive release results obtained from MELCOR severe accident analyses. Source term results for representative STC accident sequences of the FRAMATOME, OPR1000, and APR1400 reactor types were used to represent the different reactor designs at the Hanul site. Fig. 2 shows the time-dependent cumulative Cs and I release fractions for the STCs of the three reactor types.

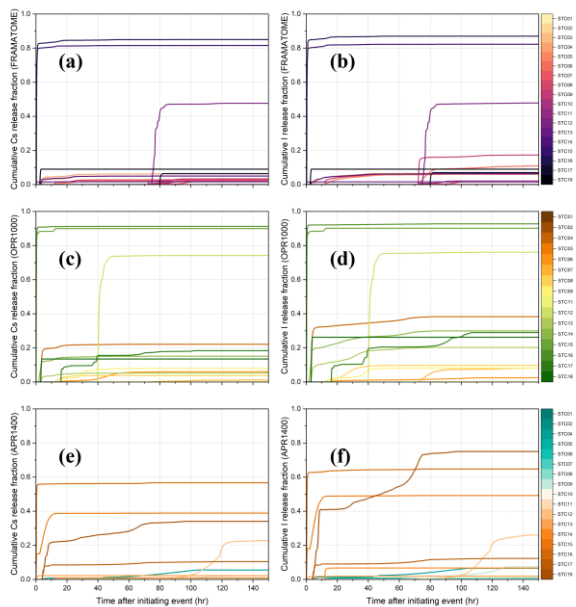


Fig. 2. Time-dependent cumulative Cs and I release fractions obtained from MELCOR analyses for the STCs of (a, b) FRAMATOME, (c, d) OPR1000, and (e, f) APR1400.

For each canonical combination, the site-level cumulative release was constructed by combining the corresponding reactor-type- and STC-specific release histories. This treatment preserves reactor-type-specific source term behavior while providing a compact representation of equivalent multi-unit combinations.

The feature vector for clustering was designed to characterize release magnitude, radionuclide composition, and time-dependent release behavior. The final feature set consisted of final Cs and I release, the Cs/I release ratio, the Cs release timing center (t_{50}), the Cs 10–90% release duration (D_{10-90}) and the late Cs release fraction after 72 h from accident initiation. The temporal diversity shown in Fig. 2 supports the inclusion of timing-related features in addition to final release magnitude.

Logarithmic transformations were applied to the Cs and I release magnitudes, the Cs/I ratio, t_{50} , and D_{10-90} , while the late-release fraction was retained as a bounded fraction. Final Cs and I releases characterize the overall release magnitude, while the Cs/I ratio represents differences in radionuclide composition. The t_{50} , D_{10-90} , and late-release fraction characterize the timing, duration, and delayed component of the release, respectively.

2.3 Clustering and Selection of Representative Combinations

The transformed feature vectors were standardized prior to clustering to minimize the effect of differences in feature scale. K-medoids clustering was performed using the Partitioning Around Medoids (PAM) algorithm. Unlike a centroid-based method, K-medoids selects an existing data point as the center of each cluster, allowing representative cases to be selected directly from the analyzed STC combinations.

The number of clusters was determined by considering internal clustering validation metrics together with cluster interpretability and the practical degree of reduction. The medoid of each cluster was defined as the central representative combination. In addition, the combination with the largest final Cs release within each cluster was identified as a maximum-Cs candidate for subsequent comparison. Combination frequency was retained for probabilistic interpretation but was not included as a clustering feature.

3. Results and Discussion

3.1 Source Term Feature Characteristics

Pearson and Spearman correlation analyses were performed to examine relationships among the selected source term features. As shown in Fig. 3, the final Cs and I release magnitudes were strongly correlated, with Pearson and Spearman coefficients of 0.96 and 0.95, respectively. In contrast, the release-magnitude features showed negative correlations with timing-related

features, particularly the Cs release timing center (t_{50}), indicating that combinations that larger-release combinations tended to exhibit smaller t_{50} values.

Positive correlations were also observed among the timing-related features, particularly between the 10–90% release duration and the late release fraction after 72 h. The Cs/I release ratio showed relatively weaker correlations with the other features, providing additional information on radionuclide composition.

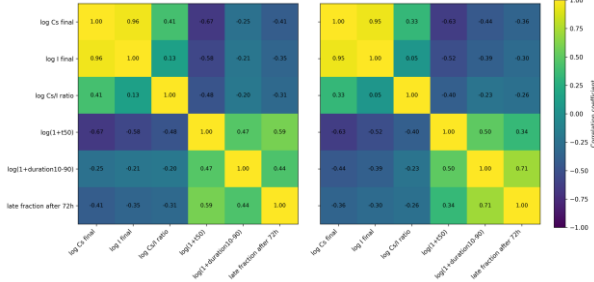


Fig. 3. Correlation matrices of the selected source term features: Pearson (left) and Spearman (right) correlations.

These results indicate that, although some features are correlated, the selected feature set retains distinct information on release magnitude, radionuclide composition, and time-dependent release behavior. The correlation analysis was therefore used to examine redundancy and complementary information among the selected features rather than as a sole criterion for feature elimination.

3.2 Clustering and Combinatorial Reduction

K-medoids/PAM clustering was performed for multiple candidate values of k , and the resulting cluster structures were evaluated using the silhouette score, Calinski–Harabasz (CH) index, and within-cluster inertia. No single validation metric identified a unique optimum. For the case study, $k=18$ was adopted by considering the validation metrics together with cluster interpretability and the practical degree of reduction, yielding a silhouette score of 0.3107 and a CH index of 2462.9. The selected k is specific to the case-study dataset, source term feature space, and clustering criteria and should not be interpreted as a universally optimal number of clusters.

The 7,832 canonical STC combinations were consequently grouped into 18 clusters according to their similarities in the standardized source term feature space. Table I summarizes the progressive reduction from the MULOOP core-damage sequences to the final case-specific clustering. Fig. 4 summarizes the mean standardized feature profiles of the 18 clusters together with their frequency contributions. The variation across the feature profiles indicates that the clusters are differentiated not only by release magnitude but also by composition, release timing, duration, and late-release behavior.

Table I: Summary of combinatorial reduction for the ten-unit Hanul MULOOP case-study model

Analysis stage	Number
MULOOP core-damage sequences	81,941
Raw STC combination records	9,458,535
Canonical STC combinations	7,832
Selected number of clusters, k	18*

*The selected k is specific to the case study dataset, source term feature space, and clustering evaluation criteria.

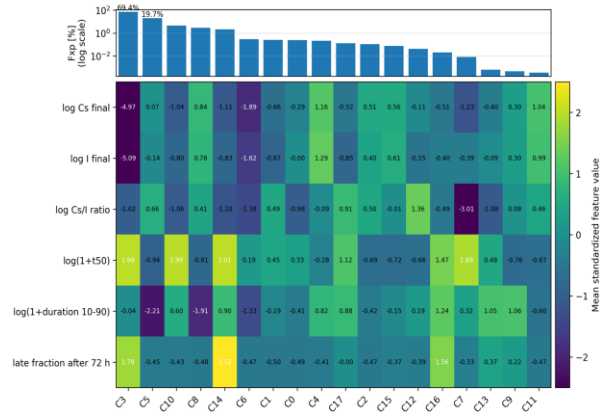


Fig. 4. Mean standardized source term feature profiles and frequency contributions of the 18 clusters.

3.3 Representative Combinations and Cluster Release Characteristics

For each cluster, the medoid and maximum-Cs candidate were identified according to the criteria described in Section 2.3. Fig. 5 compares their cumulative Cs release profiles with the pointwise P50 and P5–P95 range of each cluster. Distinct temporal patterns were observed among the clusters. For example, C3(Cluster 3) had the largest frequency contribution but exhibited a relatively low Cs release, whereas C5 developed a comparatively large cumulative release at an early stage. C8 showed a sharp initial increase followed by a plateau, while C7 exhibited a pronounced delayed increase approximately 70–110 h after the initiating event. These results illustrate that the clustering distinguishes temporal release patterns that cannot be described by final release magnitude alone.

Frequency contribution and source term severity were not necessarily aligned. Clusters with relatively high frequency contributions did not always exhibit the largest Cs releases, while some lower-frequency clusters showed larger releases or distinct temporal behavior. Selecting cases based solely on frequency or a single release-magnitude criterion may therefore overlook combinations with distinct source term characteristics.

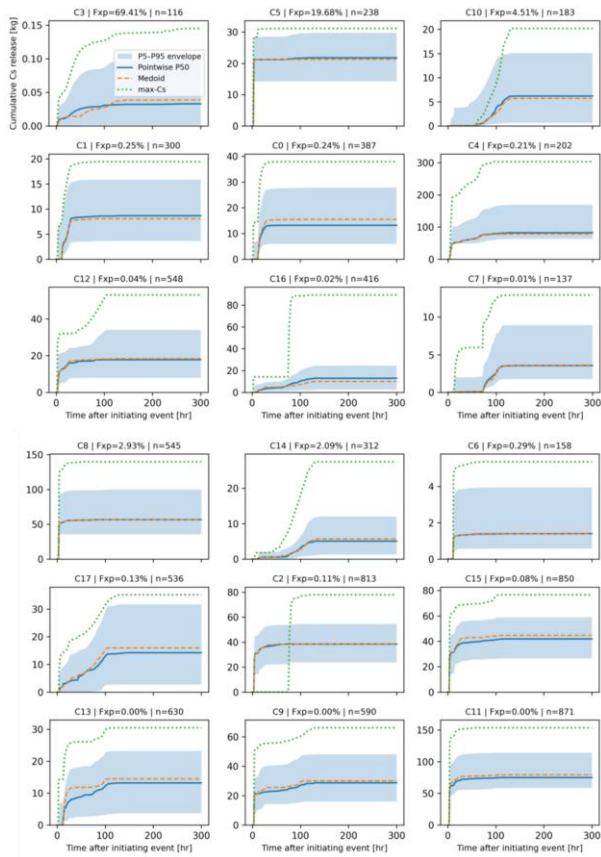


Fig. 5. Cluster-wise cumulative Cs release profiles for the selected $k=18$, including the pointwise P5–P95 envelope, P50, medoid, and maximum-Cs profiles. Clusters are ordered by their total frequency contribution.

The medoid representatives and maximum-Cs candidates identified in this study should be interpreted as source term inputs for subsequent consequence comparison, rather than as cases already demonstrated to reproduce or bound offsite consequences. In particular, the maximum-Cs candidate represents a high-release case only with respect to final Cs release and should not be interpreted as a universal bounding case for all consequence endpoints. In addition, canonical aggregation was applied only for source term characterization; unit-specific assignments and physical release locations should be reintroduced when spatial effects are explicitly considered in subsequent consequence analysis. Level 3 consequence analyses of the selected medoid representatives, maximum-Cs candidates, and additional comparison cases are currently underway to evaluate their consequence-level representativeness and conservatism.

4. Conclusions

A combinatorial reduction methodology based on K-medoids clustering was applied to multi-unit PSA by integrating Level 1 and Level 2 PSA results with source term information obtained from MELCOR analyses. For the ten-unit Hanul site MULOOP case study, approximately 9.46 million raw STC combination

records were generated, from which 7,832 canonical combinations were retained after aggregation and screening for source term data availability. The canonical combinations were characterized using release magnitude, radionuclide composition, and time-dependent release behavior and were subsequently organized into 18 case-specific clusters. For each cluster, the medoid was selected as the central representative, while the maximum-Cs combination was identified as a high-release candidate for subsequent consequence comparison.

The results show that a large multi-unit source term combination space can be systematically reduced by considering multiple source term characteristics rather than relying solely on accident frequency or a single release-magnitude criterion. The selected number of clusters and representative combinations are specific to the analyzed case, feature space, and clustering criteria. Level 3 consequence analyses of the selected medoid representatives, maximum-Cs candidates, and additional comparison cases are currently underway to evaluate their consequence-level representativeness and conservatism and to validate the proposed reduction methodology beyond the source term feature space. Future work will further extend the methodology to multi-unit accident scenarios initiated by external hazards, such as seismic events, and explore its applicability to multi-module SMRs.

Acknowledgements

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 (RS-2025-02643022).

REFERENCES

- [1] International Atomic Energy Agency (IAEA), Multi-unit Probabilistic Safety Assessment, Safety Reports Series No. 110, 2023.
- [2] N. E. Bixler and S. Kim, Performing a Multi-Unit Level-3 PSA with MACCS, Nuclear Engineering and Technology, Vol. 53, pp. 386-392, 2021.
- [3] W. Song, S. Park, Y. Seo, and M. Jae, A Source Term Binning Methodology for Multi-Unit Consequence Analyses, Reliability Engineering & System Safety, Vol. 202, 106989, 2020.
- [4] M. Cho, D. Lim, C. Park, and M. Jae, Development of a Site Release Category Model and Its Application, Annals of Nuclear Energy, Vol. 196, 110196, 2024.
- [5] K. Jin, J. Cho, and S. Kim, Machine Learning-Based Categorization of Source Terms for Risk Assessment of Nuclear Power Plants, Nuclear Engineering and Technology, Vol. 54, pp. 3336-3346, 2022.
- [6] J. Cho, S. Han, D. Kim, and H. Lim, Multi-Unit Level 2 Probabilistic Safety Assessment: Approaches and Their Application to a Six-Unit Nuclear Power Plant Site, Nuclear Engineering and Technology, Vol. 50, pp. 1234-1245, 2018.