Comparison of Conventional Scaling Factor Methods and Random Forest for Radionuclide Inventory Evaluation of Activated Metal Waste

Yeeun Choi[a], Jong-Yun Kim[b,c], Tae-Hyeong Kim[b], Jeongmook Lee[b,c], Sang Ho Lim[b,c], Nak-Woong Yang[d], Chang-Je Park[d], Jeongbeen Lee[a], Donggyu Lee[a], Hongjun Park[a]*
[a]*Department of Nuclear and Energy Systems Engineering (Nuclear Engineering), Dongguk University WISE Campus, Gyeongju, Republic of Korea*
[b]*Korea Atomic Energy Research Institute, 111, Daedeok-daero 989beon-gil, Yuseong-gu, Daejeon, Republic of Korea*
[c]*University of Science and Technology, 217, Gajeong-ro, Yuseong-gu, Daejeon, Republic of Korea*
[d]*Department of Quantum and Nuclear Engineering, Sejong University, Neungdong-ro, Gwangjin-gu, Seoul, Republic of Korea*
* *Corresponding author: hjpark86@dongguk.ac.kr*

***Keywords :** Radioactive Waste, Radionuclide Inventory, Scaling Factor, Activated Metal Waste, Random Forest

1. Introduction

Radionuclide inventories of difficult-to-measure (DTM) radionuclides, such as Fe-55 and Ni-63, are difficult to directly measure in radioactive waste. Therefore, the scaling factor (SF) method has been widely used to estimate DTM radionuclide inventories based on their relationships with easily measurable key nuclides (KNs). Conventional SF estimation methods are generally based on activity ratios or regression relationships between key and DTM radionuclides [1,2]. Recent simulation-based studies on activated decommissioning waste have shown that the relationships of Fe-55 and Ni-63 with Co-60 can exhibit nonlinear behavior as activation levels and neutron flux conditions vary [3]. Random Forest (RF) can model such nonlinear relationships and incorporate multiple input variables without assuming a predefined functional form [4,5]. In this study, ORIGEN-generated activation data for SUS304 were used to compare the predictive performance of the geometric mean, log-linear regression, RF-1, and RF-2 for Fe-55 and Ni-63. RF-1 used Co-60 activity alone, whereas RF-2 additionally included neutron flux and Co concentration.

2. Materials and Methods

2.1 ORIGEN calculation dataset

The dataset consisted of 30 ORIGEN calculation cases for activated SUS304 metal waste. Co-60 was selected as the key nuclide, and Fe-55 and Ni-63 were selected as the target DTM radionuclides. Activities at a cooling time of 20 years were used for all radionuclides. The Co impurity concentration ranged from 449 to 2,501 ppm, and the neutron flux was varied across LOW, MID, and HIGH conditions, resulting in a final range of 1.10×10^{11}~7.73×10^{12} n/cm²·s. To maintain the total material mass balance with changes in Co concentration, the Fe composition was adjusted while the Ni composition was kept constant.

2.2 Prediction modeling and cross-validation

Four methods were compared: geometric mean (GM), log-linear regression (LLR), RF-1, and RF-2. For GM, the geometric mean of the DTM-to-Co-60 activity ratios in each training set was used as the representative SF. LLR and RF-1 used only Co-60 activity as the input variable, whereas RF-2 additionally included neutron flux and Co concentration. All methods were evaluated using the same 5-fold cross-validation scheme, and predictive performance was calculated for cases not included in the corresponding training fold.

2.3 Performance metrics

Model performance was evaluated using R^2, RMSE, MAE, RMSE($\log_{10}$), and Factor-of-2 (Fo2). Higher R^2 values indicate better agreement with the variation in the ORIGEN-calculated activities, whereas lower RMSE and MAE values indicate smaller absolute prediction errors in Bq. RMSE($\log_{10}$) evaluates prediction errors on a logarithmic scale, while Fo2 represents the percentage of predictions within a factor of two (0.5–2.0) of the ORIGEN-calculated activities. These metrics were considered together because they characterize different aspects of prediction performance.

3. Results and Discussion

The 5-fold cross-validation results for the 30 cases are summarized in Table I. When only Co-60 activity was used as the input, LLR showed higher R^2 values and lower RMSE, MAE, and RMSE($\log_{10}$) values than RF-1 for both radionuclides. Thus, for the present dataset, LLR provided better predictive performance than RF-1 under the same single-input condition.

Table I. Prediction performance of the four methods

Method	R^2	RMSE (Bq)	MAE (Bq)	RMSE ($\log_{10}$)	Fo2 (%)
Fe-55					
GM	0.632	7.73×10^5	4.51×10^5	0.225	83.3
LLR	0.871	4.57×10^5	3.10×10^5	0.211	80.0
RF-1	0.833	5.21×10^5	3.75×10^5	0.276	73.3
RF-2	0.960	2.55×10^5	1.68×10^5	0.108	100.0
Ni-63					
GM	0.623	9.24×10^6	5.36×10^6	0.225	83.3
LLR	0.872	5.39×10^6	3.65×10^6	0.211	80.0
RF-1	0.834	6.14×10^6	4.43×10^6	0.275	73.3
RF-2	0.961	2.99×10^6	1.98×10^6	0.107	100.0

RF-2 showed the highest R^2 values and the lowest prediction errors for both radionuclides, with an Fo2 of 100% in both cases. Compared with LLR, RF-2 reduced RMSE by approximately 44–45%, MAE by approximately 46%, and RMSE($\log_{10}$) by approximately 49%, indicating improved performance in both absolute and logarithmic error metrics.

LLR outperformed GM in R^2, RMSE, MAE, and RMSE($\log_{10}$); however, the Fo2 values were 83.3% for GM and 80.0% for LLR. This difference arises because Fo2 only evaluates whether a prediction falls within the 0.5–2.0 range of the ORIGEN-calculated activity and does not distinguish the magnitude of errors within that range. Therefore, Fo2 should be interpreted together with continuous error metrics rather than used as a standalone performance measure.

Unlike LLR and RF-1, which used only Co-60 activity as the input, RF-2 additionally included neutron flux and Co concentration. Therefore, the improved performance of RF-2 should be interpreted as the combined effect of the RF model and the additional input variables rather than as an improvement resulting solely from the use of RF. Because the present results were obtained within the specified SUS304 composition and activation conditions, further validation using independent calculation conditions and measured data is required.

4. Conclusions

In this study, the predictive performance of GM, LLR, RF-1, and RF-2 was compared using ORIGEN-generated activation data for SUS304. Under the single-input condition using only Co-60 activity, LLR outperformed RF-1, whereas RF-2, which additionally included neutron flux and Co concentration, showed the best predictive performance for both Fe-55 and Ni-63. Differences in model rankings among the performance metrics indicate that model performance should be assessed using multiple error metrics rather than a single metric. Further validation with expanded datasets and independent measured data is required, together with assessment of the practical availability and discriminatory value of input variables in waste characterization.

REFERENCES

[1] International Atomic Energy Agency, Determination and Use of Scaling Factors for Waste Characterization in Nuclear Power Plants, IAEA Nuclear Energy Series No. NW-T-1.18, Vienna, 2009.

[2] International Organization for Standardization, Nuclear Energy - Nuclear Fuel Technology - Scaling Factor Method to Determine the Radioactivity of Low- and Intermediate-Level Radioactive Waste Packages Generated at Nuclear Power Plants, ISO 21238:2007, Geneva, 2007.

[3] T.-H. Kim, B. Kang, D. W. Lee, S. H. Lim, J. Kim, S. G. Cho, C.-J. Park, J.-Y. Kim, and J. Lee, A Simulation-Based Framework for Scaling Factor Modeling and GMM-Based Zonal Stratification in Activated Decommissioning Waste, Nuclear Engineering and Technology, Vol.58, No.2, 103981, 2026.

[4] L. Breiman, Random Forests, Machine Learning, Vol.45, pp.5-32, 2001.

[5] B. Zaffora, J.-P. Chevalier, F. P. La Torre, C. Luccioni, M. Magistris, and G. Saporta, Statistical Learning Approaches Applied to the Calculation of Scaling Factors for Radioactive Waste Characterization, Proceedings of the 14th Congress of the International Radiation Protection Association (IRPA14), May 9-13, 2016, Cape Town, South Africa.