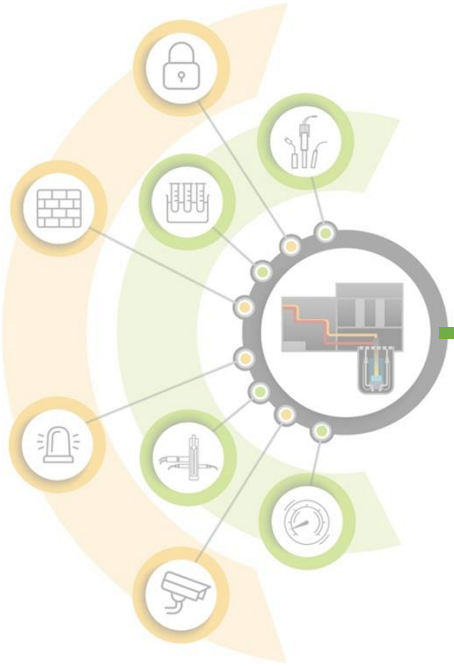

Advanced Reactor Safeguards

Use Machine Learning to Improve Burnup Measurement in Pebble Bed Reactors

FY23 Annual Report

Prepared for
US Department of Energy

Yonggang Cui, Carlos Soto, Odera Dim
Brookhaven National Laboratory

September 2023

BNL-NN-20230717-0185-00-FORE

ABSTRACT

Burnup measurement is an important step in material control and accountancy (MC&A) at nuclear reactors. BNL has been investigating using machine learning (ML) method to improve the accuracy of burnup measurements, specifically for fuel coming out of pebble bed reactors (PBRs). In the early phases of the project, we developed a simple PBR burnup model and used it to generate synthetic gamma-ray spectra datasets. In parallel, we developed a multilayer perceptron (MLP) machine learning regression model for burnup prediction. The model outperformed the conventional linear regression model in our tests. In FY2023, we further improved the burnup model and completed a full-core PBR simulation model so we could validate the simulation with the existing data of well-studied PBR reactors, i.e., PBMR400. In parallel, we are also looking into the explainability of the MLP regression model. This annual report details what we have accomplished in these two tasks this year.

ACKNOWLEDGEMENTS

The work is funded through the Office of Nuclear Energy, Advanced Reactor Safeguards Program, within the United States Department of Energy.

CONTENTS

This page left blank

ACRONYMS AND DEFINITIONS

Abbreviation	Definition
ARS	Advanced Reactor Safeguards
CNN	Convolutional neural network
DOE	Department of Energy
LIME	Local Interpretable Model-Agnostic Explanations
LWR	Light-water Reactor
MAPE	Mean Average Percentage Error
MC&A	Material Control and Accountancy
ML	Machine Learning
MLP	Multilayer Perceptron
R^2	Coefficient of Determination (square of the Pearson correlation coefficient)
IAEA	International Atomic Energy Agency
ID	Inventory Difference

1. INTRODUCTION

Advanced pebble bed reactor (PBR) designs pose new challenges in material control and accountancy (MC&A) because the fuel materials, distributed in many discrete pebbles, are continuously circulated through the reactor core and the refueling path compared to the bulk fuel assembly design in conventional reactors, e.g., light water reactors. PBRs are fueled with hundreds of thousands of fuel pebbles. During the normal operation of a PBR, ejected pebbles are returned to the reactor or discharged depending on the fuel burnup and physical condition.

The burnup measurement is usually based on detection of radiation signatures of fission products. Years of research has shown that measurements of fission products, such as ^{134}Cs, ^{137}Cs, ^{154}Eu, etc., can be applied independently or in combination to infer or predict the level of burnup in the fuel (Akyurek, Tucker, & Usman, 2014). A simple criterion for selecting an isotope for burnup indication is the exhibition of a strong gamma photopeak. However, it remains challenging to measure this complex source due to self-shielding effects, strong radiation background and intervening materials. Another challenge is the required high throughput in the burnup measurements. Accommodating this throughput necessitates limited measurement time and thus impacts efficiency of this measurement. A high-performing spectral analysis method is therefore required to identify patterns swiftly and accurately in the time-constrained gamma spectrum measurements.

In this study, BNL develops machine learning (ML) methods to interpret gamma-ray spectra and predict the burnup values of the pebbles. ML has achieved widespread success and adoption across numerous domains that require pattern recognition and analysis in varied data types (Butler, 2018) (Carleo, 2019). Modern deep learning approaches have supplanted hand-crafted features by learning entirely novel, yet meaningful, features and data representations directly from the raw data via deep neural network architectures; this has led to state-of-the-art and even superhuman performance on a broad range of detection, interpretive, and analytical tasks. In the first two years of this project, we have developed multilayer perceptron (MLPs) and convolutional neural networks (CNNs) for burnup prediction. To support the ML algorithm development, we developed a modeling and simulation workflow to simulate the burnup of a single pebble and detector response in spectroscopic measurements. Our latest results have demonstrated that ML is able to improve the prediction accuracy by a factor of 4 compared to the single or multiple photopeaks based linear regression method.

In 2023, we focus on validating the simulation model and investigating the explainability of the ML models in burnup prediction tasks. This report details the research results this year.

2. MODELING AND SIMULATION

2.1. Addition of Collimator Model to Transportation Simulation

In our early research, all data related to burnt pebbles was generated with the assumption that the ejected pebble is measured with a detector. In real operation, pebbles ejected from a PBR core are highly radioactive and can only be measured behind a collimator. To model the photon transportation from the pebble surface to the detector surface more accurately, we updated our Serpent model to include a Lead collimator. Although we have the option of varying the thickness and aperture of the collimator, we arrived at an optimal thickness and aperture size of 8 cm and 1 cm respectively.

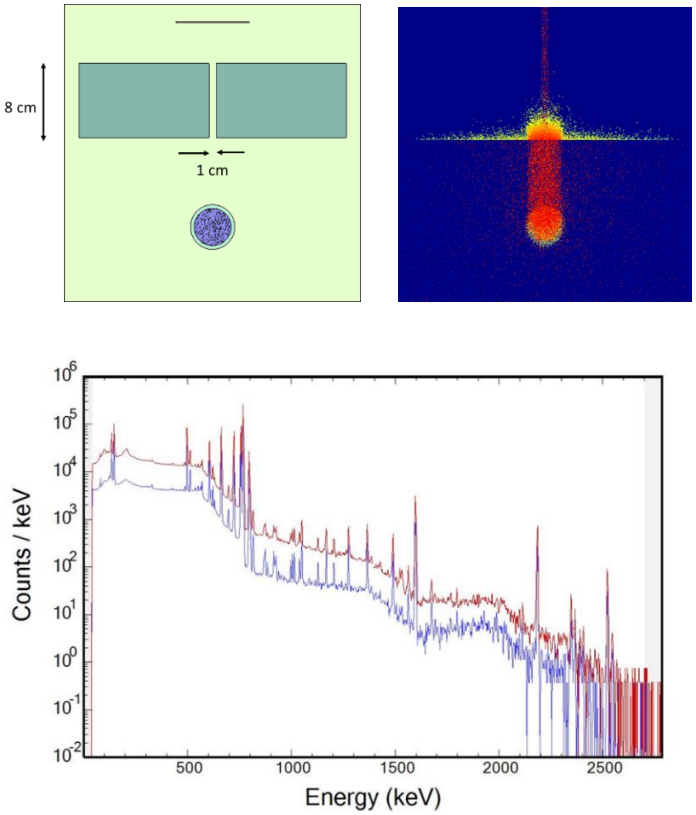

Figure 1 Modeling collimator in Serpent. Top left: basic collimator model. Dark green is collimator material, e.g., lead or tungsten. Light green is free space. Purple is the fuel pebble. Top right: simulation of photo flux before and after the collimator. Bottom: gamma-ray spectra of a fuel pebble with and without a collimator

Understanding that the inclusion of collimation to our model might introduce some reduction in the efficiency or performance of the simulation. We also studied the various implementations of weighting windows to combat this. The two approaches studied were the basic and adaptive approaches. The basic approach works by creating a mesh over the full geometry and assigning importance to each. Then during the transport process, it preferentially tracks particles that lead to target location (i.e., radiation detector). The adaptive approach is slightly different and is better suited to the complex shielding problem. The adaptive approach adds an extra step of dynamically

creating new importance when there are material density changes encountered by a radiation particle along its track.

Weighting windows are not currently in use as the performance of the simulation is not hampered by the inclusion of collimation as much in our current configuration as we still achieve low uncertainties.

2.2. Independent Transport and Burnup

The initial model used in generating the spectra applied in the ML task was done with the aim of efficiently generating the data in an automated fashion. This process allows for large sets of data with varying sensitivity of a few parameters to be produced in a relatively short amount of time. However, a few situations arise that make a combined burnup and transport model combined to be relatively challenging. A few of these scenarios include:

1. Verifying the output from a single burnup level with a variety of transport setups or configurations.
2. Testing some of the aspects of either the burnup or transport in a timely fashion.

For these reasons and others, the automation model was split into an independent burnup and transport models. To perform this split we re-wrote the Python script that automates the generation of the Serpent burnup model to output a binary file that contains the composition and other parameters from each burn step. A second python script was written that requires the binary output file from the initial burnup simulation to generate a source term form the transport simulation. The script also supports using a binary file containing multiple burn steps to generate multiple source terms for a given transport configuration.

2.3. Model of Full PBR Reactor Core

Earlier approaches to generating data sets for this project were based on a simplified 3x3x3 lattice model. Such models are useful to give a generic perspective of what the burnup and transient isotopic composition would be. However, these simplistic models do not represent any true reactor system which means the flux and power profiles as well as burnup levels attained in the reference pebbles for the lattice model are simply theoretical. For this reason, we attempted to create a full core model of a well-documented PBR reactor design (i.e., PBMR). The model can be used in two ways, either as an independent model that directly produces data for the ML algorithm or one to specify operational parameters for the lattice model which then generates the data for ML.

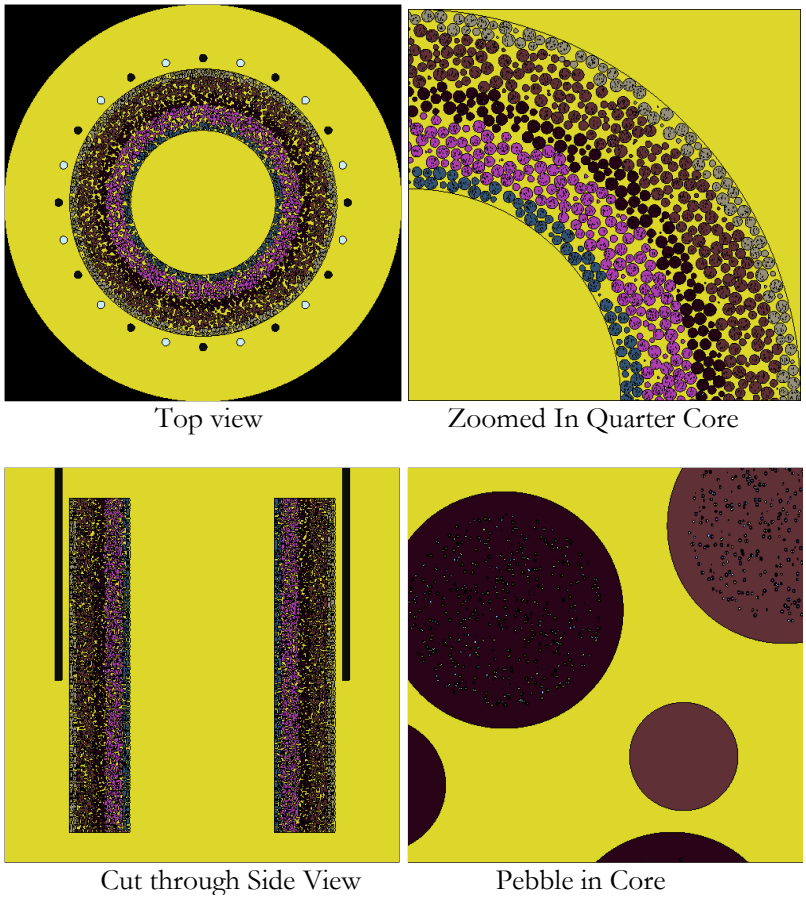

Figure 2 Full-core PBR simulation model with 5 zones

2.4. Comparisons between Serpent and ORIGEN Simulation Results

To develop and validate the performance of the full core model, we compared the Serpent and Origen models. Three reactor metrics were compared: specific isotopes, the neutron and gamma emission rates.

A total of 13 depletion steps and 12 decay steps were used for the simulation. The cooling steps of the simulation were defined as decay only to save the unnecessary time used by Serpent in performing transport after each cooling step. Five special pebbles, which would be in each of the five artificially designated but commonly used radial channels of the core, were selected. The boundaries of these five radial channels were defined as 100, 109, 135, 150, 176 and 185 cm radii. The simulation in Serpent was performed to achieve a core average burnup of about 90 GWd/MTU. When this is done, the average neutron spectra and power in each of the five pebbles were extracted from the Serpent results and input to ORIGEN. The 252-group neutron spectra from Serpent were then used by ORIGEN to produce the one-group library needed for depletion and decay calculations. This approach aimed to verify Serpent's isotopic transmutations and decay results. This approach also eliminated the need to build a 3D core model in SCALE and is consistent with the Serpent model and the potential discrepancies in neutron spectra between the Serpent and SCALE results, although 3D SCALE models exist for PBMR-400. The isotopic concentrations of a few nuclides of interest and neutron and gamma source terms produced from the ORIGEN calculations were compared with those from the Serpent calculations.

There are two stages of simulation in Serpent; the burnup and decay phase is where the transient compositions, spatial and local power as well as 252 group fluxes are tallied. The second phase uses the reactor state at the steps corresponding to cooling times in Table 2 to start a source transport simulation which is used to determine the 25-group gamma and 21-group neutron emission spectra after the specified cooling period elapses. Both phases of the Serpent simulation were performed with about 5 million neutron histories (this is possible because Serpent provides functionality to resume a simulation based on a previous run using pre-generated binary files). The highest uncertainty in all results was recorded in the pebble fluxes. The bin uncertainties were on the order of 10 – 20 %.

Table 1 Burnup simulation results with the full-core PBR model

Pebble	Volume (cm3)	Burnup (MWD/kgU)	Initial density of U (g/cm3)	Initial mass of U (gU)	Initial mass of U (kgU)	Duration of Burnup (days)	Power (MW)	Specific power (MW/KgU)
1	9.8E-01	6.98E+01	9.20	9.00	9.00E-03	4.66E+02	1.35E-03	1.50E-01
2	9.8E-01	6.20E+01	9.20	9.00	9.00E-03	4.66E+02	1.18E-03	1.33E-01
3	9.8E-01	5.15E+01	9.20	9.00	9.00E-03	4.66E+02	9.76E-04	1.10E-01
4	9.8E-01	3.86E+01	9.20	9.00	9.00E-03	4.66E+02	7.32E-04	8.29E-02
5	9.8E-01	3.33E+01	9.20	9.00	9.00E-03	4.66E+02	6.31E-04	7.14E-02

As mentioned earlier, three reactor fuel parameters were extracted for the spent fuel and compared between Serpent and Origen. These three parameters were obtained for five radials zones (i.e., a pebble in each radial zone) of the PBMR core. Figures 3-6 show the selected isotope compositions and neutron and gamma emission rates for the 2nd radial zone from the center of the reactor core.

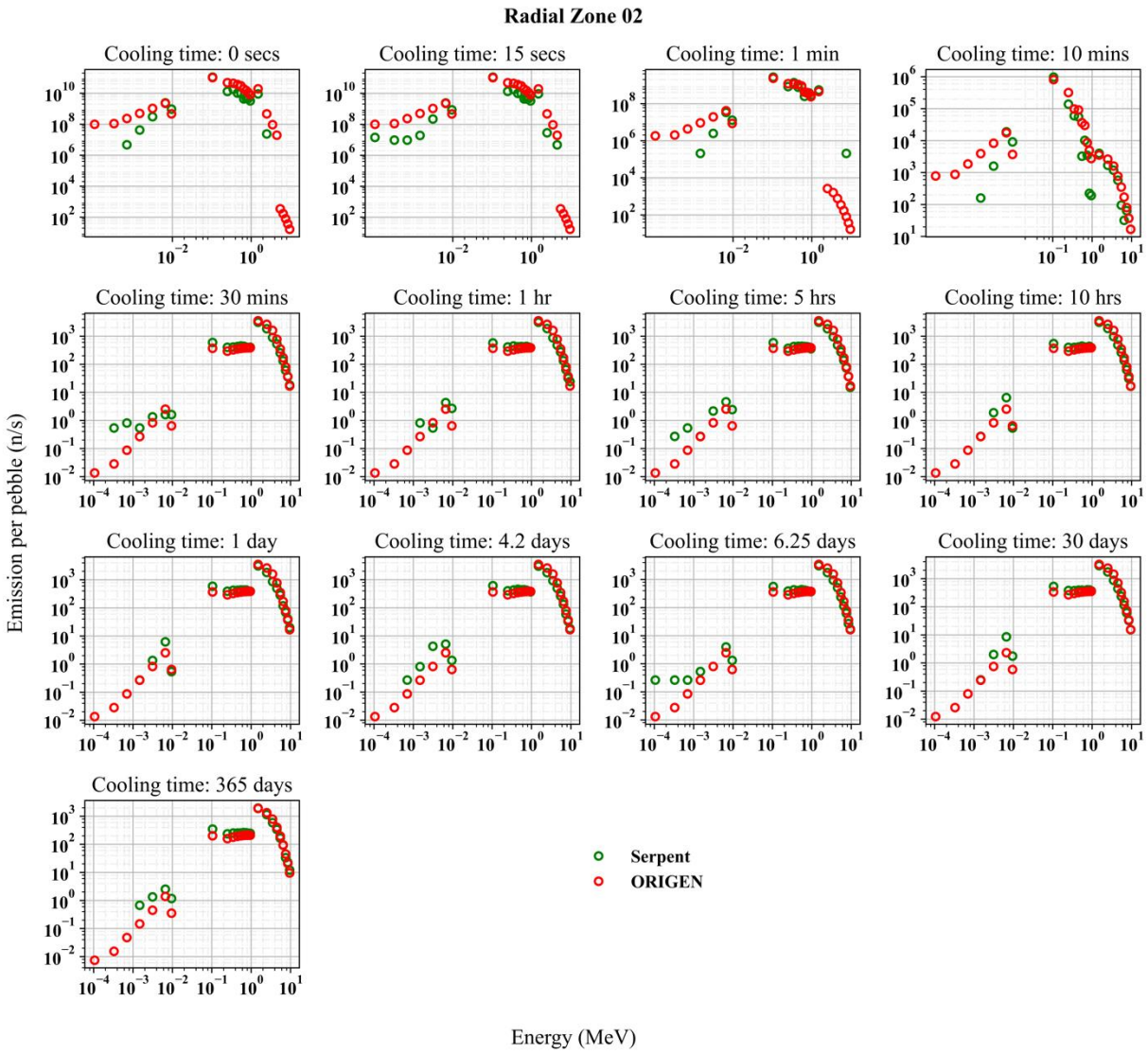

Figure 3 Comparison of the neutron emission rates.

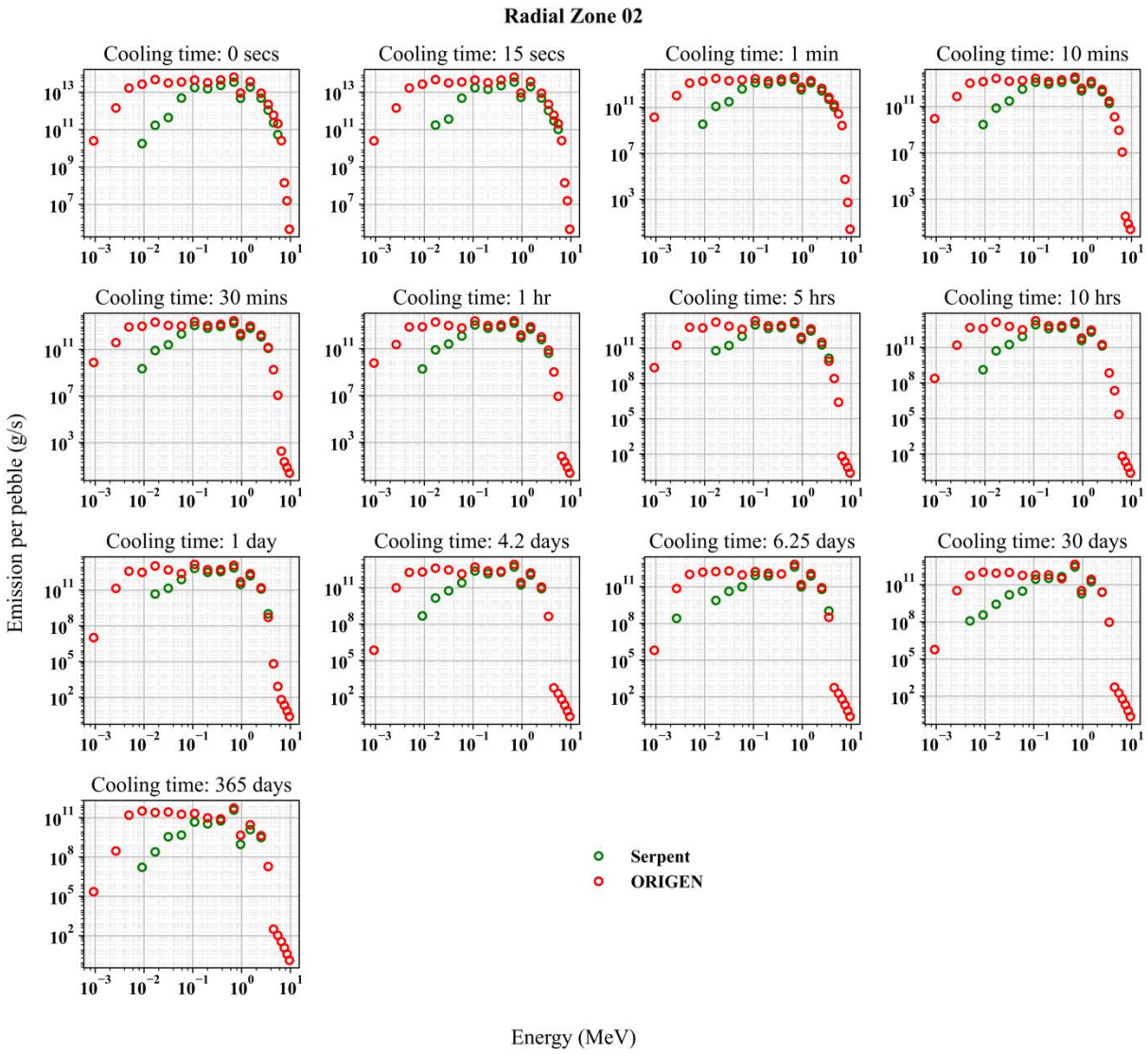

Figure 4 Comparison of the gamma emission rates.

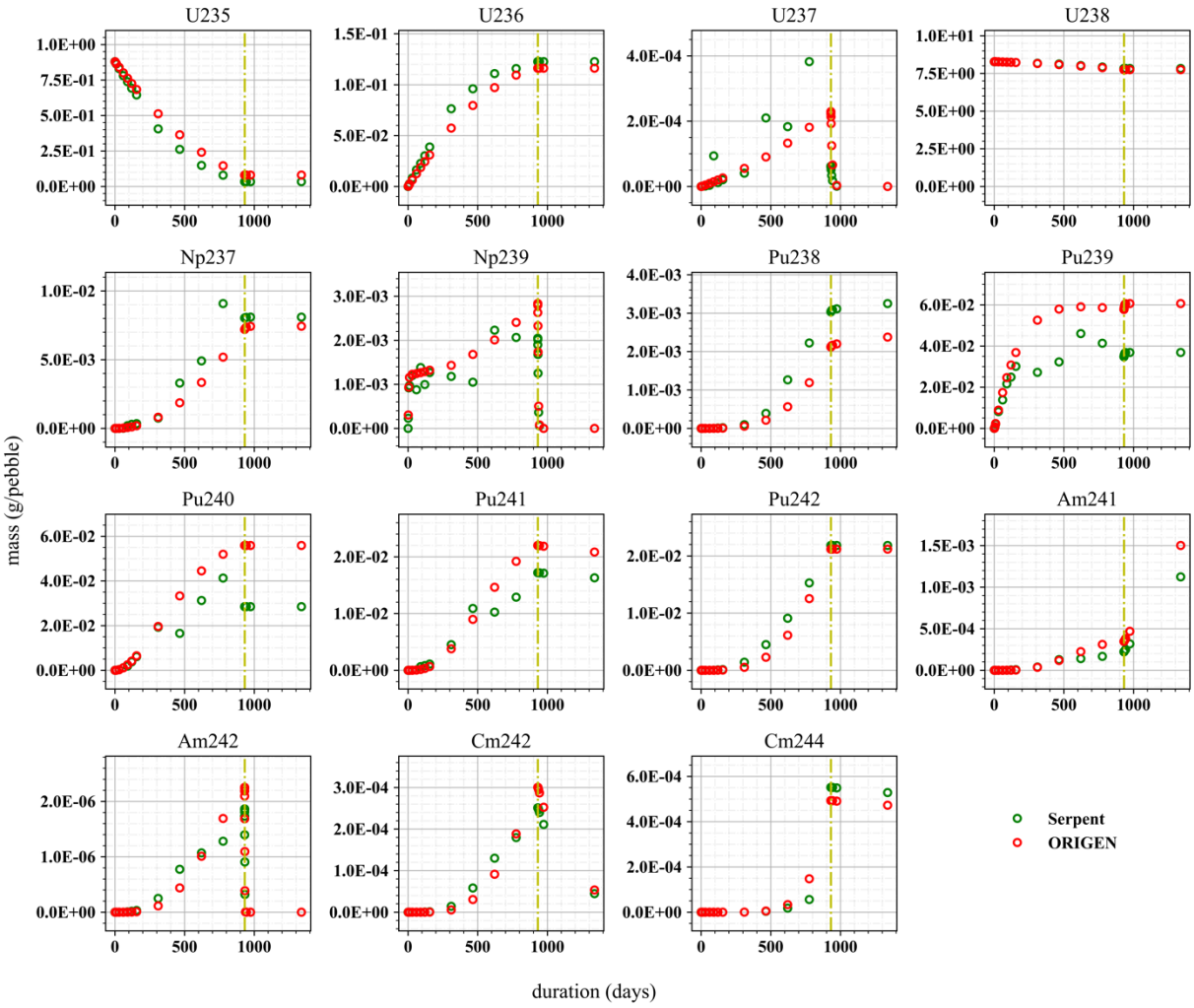

Figure 5 Comparison of the burnup composition of selected isotopes.

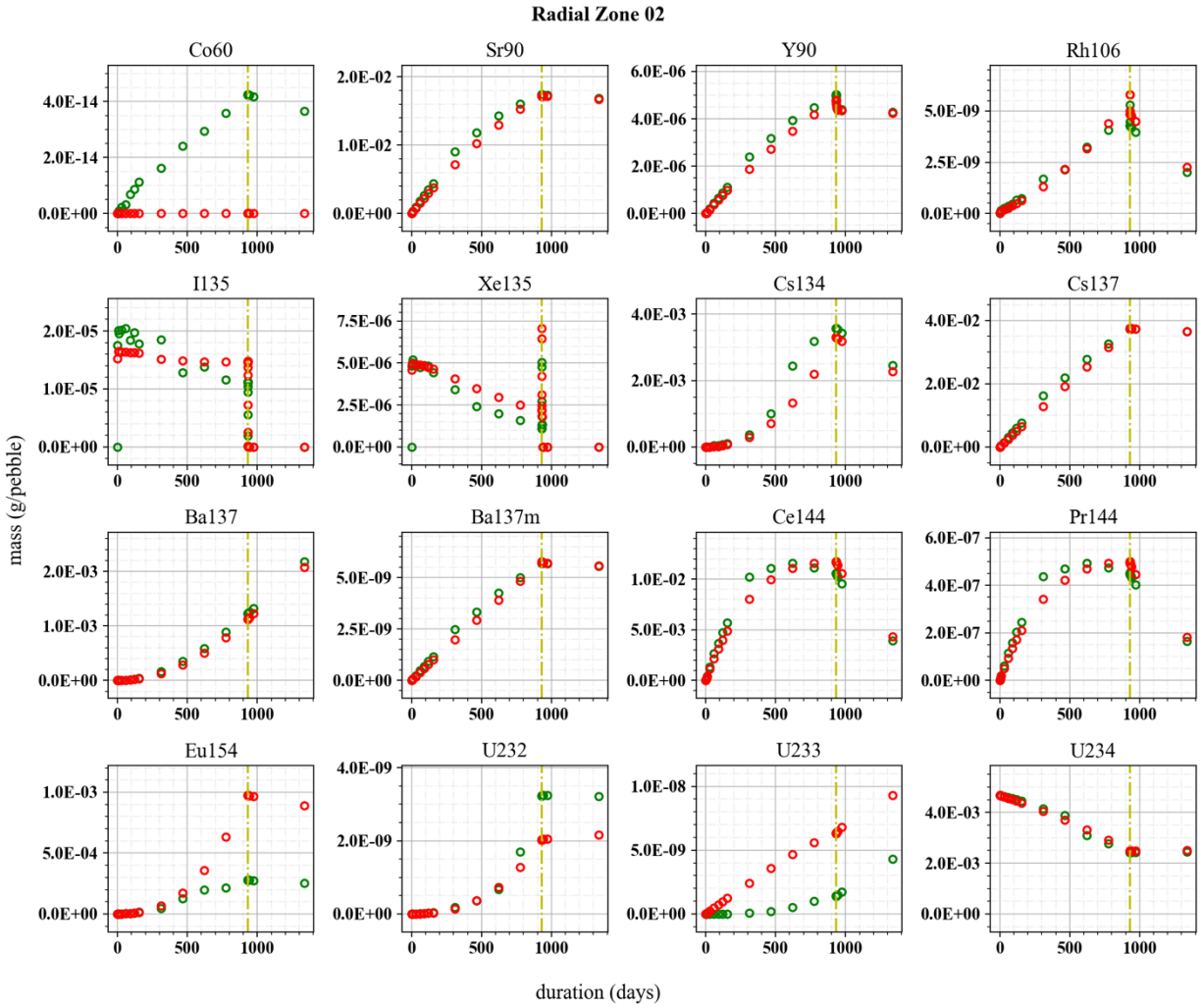

Figure 6 Comparison of the burnup composition of selected isotopes (cont'd).

3. EXPLAINABILITY OF ML MODEL

Our earlier results in this project [3] have shown that a deep learning-based ML approach can be successfully applied to improve PBR burnup prediction. However, deep learning methods produce solutions that are inherently difficult to interpret, so investigating the explainability of the models is important for the ML approach to be accepted by the community. In this section, we describe our approaches to introduce ML explainability to our methods and thus better interpret the outcome and impact.

3.1. Software Tools for this Research

We selected techniques to examine ML explainability (also called XAI, for explainable AI) based on two criteria. First, the techniques must support the data and models we previously developed. These consisted of several thousand simulated gamma spectra and ML models of three types: multi-layer perceptrons (MLPs), convolutional neural networks (CNNs), and hybrid models using CNN input layers and MLP intermediate and output layers. Second, the techniques must offer explainability insights with respect to the input data. That is, they should identify simple features and relationships in the gamma spectra that are most responsible for the ML models' predictions (as opposed to attempting to interpret the deeper abstract features that the ML models learn).

Based on these criteria, we selected LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (Shapley Additive Explanations), two open-source, model-agnostic techniques designed to examine raw feature contributions by dynamically producing surrogate inference models. The two techniques operate differently and complement one another well. LIME produces linear surrogate models for individual sample inferences by automatically perturbing input-space features and using these to train small linear models that are compared to the original ML result. SHAP, on the other hand, produces an ensemble of tree-based models by using the Shapley value concept from cooperative game theory to estimate the marginal contribution of individual features to all possible coalitions of features.

An important challenge affecting both XAI techniques is that they are best suited to small feature sets and do not scale well to hundreds or thousands of features (SHAP is particularly sensitive in this respect). For gamma spectra signals, the set of input features is the number of channels, which in our case is 4096. We are nonetheless interested in resolving precise energies in the spectra that are key contributing features for the ML methods, so we designed a tiered strategy to carry out our explainability analysis. We begin by iteratively re-binning spectra and using these to perform LIME feature selection. This involves starting at a very coarse resolution (64 channel), applying the LIME technique, filtering energy ranges according to LIME-determined relevance, re-binning at a higher resolution, and re-applying LIME until prominent features are selected from the original spectra resolution. Following this, the final features are used for SHAP analysis for comparison to LIME results.

3.2. Explainability Results

Our explainability results follow from our iterative re-binning strategy with the LIME technique to narrow down the energy bands with likely causal relationships to burnup prediction. After several iterations we settled on a 3-stage re-binning and LIME feature selection approach. The first stage re-bins the original 4096-channel spectra into 64 low-resolution bins and applies LIME to select 16 of these bins (i.e., 25% of the features). According to the LIME package's explainability metrics, this

quarter of the features captures 88.74% of the explanatory power available to the LIME approach. Note that this does not necessarily correspond to a 88.74% correlation with true burnup, nor to 84.74% of the performance of the ML method under investigation. Rather, this explanatory metric refers to the proportion of the maximum correlation possible between a linear combination of all available perturbation-based features and the ML model being explained.

The choice of feature subset cutoff determines this explainability metric, and it is therefore advantageous for the purposes of explainability that the majority of this correlation be associated with as few features as possible (as was the case in our experiments), particularly as we need to further subdivide these features in our multi-stage approach. Note also that the rationale for these correlation metrics being considered "explainability" metrics is because LIME uses feature perturbations from individual examples to generate synthetic training data that originates entirely from the input to the ML model. The correlation between a linear surrogate model built on this synthetic data and the actual ML model's prediction is therefore an attempt to explain which of the ML model's input features had an outsized impact on its prediction.

Because LIME operates only on individual prediction examples, we re-ran the LIME process for each of the ~200 examples set aside from the ML method's training data for the purpose of testing and evaluation and determined feature explainability according to the aggregate of the explainability scores for all samples in the test set. We did this for 2 measurement conditions: 12-hour cooling and 120-hour cooling (both with 1-hour acquisition time). Figure 7 shows LIME results for stage 1 in the 12-hour cooling condition.

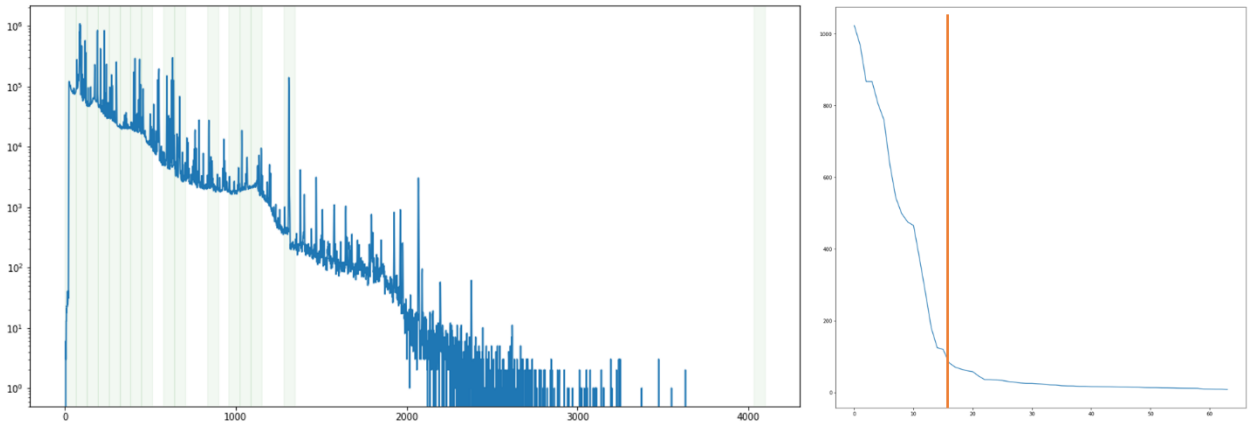

Figure 7 *Left*: The top 25% features selected by the LIME method (vertical light green bars) for the 12hr cooling dataset in the 1st stage of the 3-stage explainable feature-identification process, overlain on a representative spectra sample from this dataset. *Right*: The proportion of "explainability" due to each of the 64 re-binned features, ranked. The top 16 of 64 bins are selected for the next round, capturing 88.74% of the method's potential explainability.

Stage 2 takes the top 16 bins selected in stage 1, maps these back to the corresponding 1024 raw energy channels in the original spectra, re-bins these by a factor of 16x to 64 new re-binned features (not contiguous) and re-applies the LIME process across the test set as described above. Again, the top 16 (25%) features are selected for the next round, and in this stage, these correspond to 78.95% of the available explanatory power. Figure 8 shows these stage 2 results, again for the 12-hour fuel cooling condition.

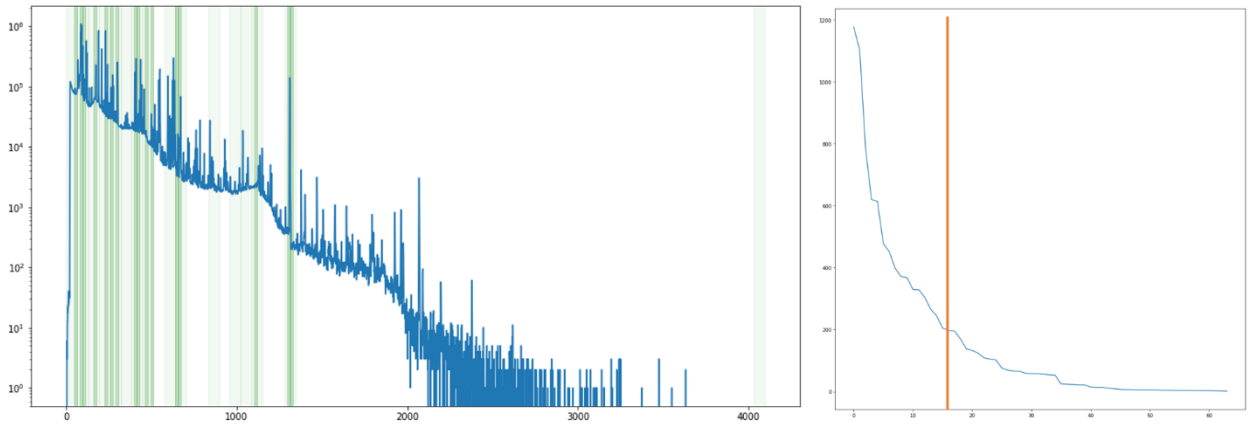

Figure 8. *Left*: Top 25% LIME features selected in stage 2 (darker vertical green bars) from the spectra subset selected in stage 1 (lighter green bars). *Right*: The top 16 of 64 features in this stage account for 78.95% of the available explainability.

In the 3rd and final LIME stage, the selected features from the previous stage were again mapped to the corresponding original energy channels (256 out of the original 4096), these were re-binned (by 4x this time) to 64 final feature candidates, and LIME was used to determine the top explainable features. As no further stages are necessary, we this time selected the top 24 features, comprising 96.71% of the LIME explainability. Figure 9 shows the location of these final features in the original energy space.

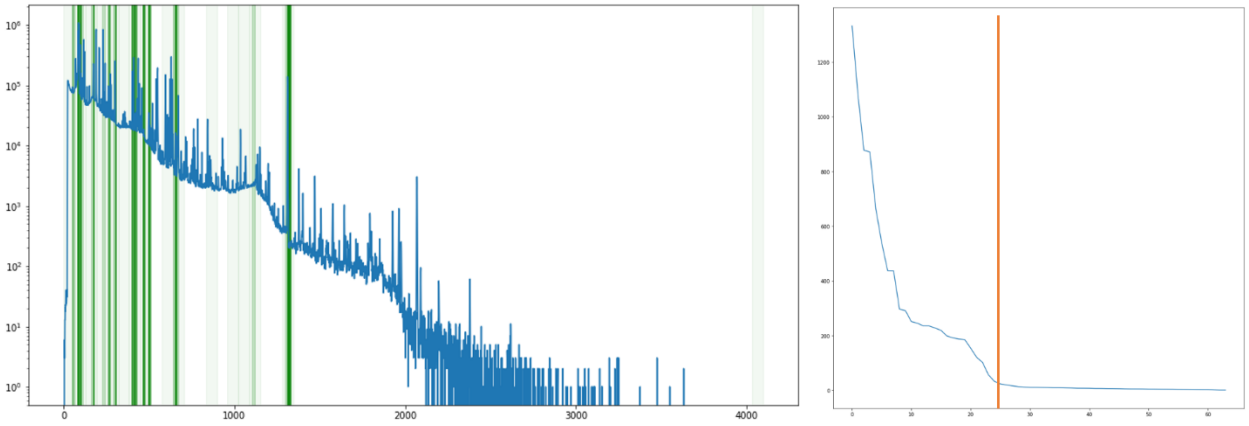

Figure 9. *Left*: Final LIME-selected features (darkest green, N=24). *Right*: The top 24 of 64 features account for 96.71% of available explainability.

Having selected the top 24 energies that best explain the source of the ML model's burnup predictions, we verified the significance of these features (as well as the intermediate features from the earlier 2 stages) by training new MLP models exclusively on these features selected from the original training data. As shown in Table 2, predictive performance of these revised models uniformly improves across all metrics after each LIME feature selection stage, strongly suggesting that our multi-stage LIME process correctly selected the dominant features used by the ML model (and thus removed the effects of noisy, uninformative features). We also compared linear regression

performance between the original spectra and the final selected features and found that the LIME-selected features also significantly improve this analysis.

Table 2. Validation of LIME-selected features with new MLP and LR models. Arrows indicate direction of better performance per metric: RMSE (root mean square error), MAPE (mean absolute percent error), R^2 (correlation coefficient).

		↓ RMSE	↓ MAPE (%)	↑ R^2
MLP	Original spectra	0.66	2.04	0.9995
	Stage 1 features	0.37	1.61	0.9998
	Stage 2 features	0.27	1.33	0.9999
	Stage 3 features	0.28	1.07	0.9999
Linear Regression	Original spectra	0.74	7.93	0.9992
	Stage 3 features	0.53	5.16	0.9996

The 24 spectra energies selected by our explainability process are shown in Table 3. We have not yet identified to which, if any, decay products these energies may correspond, but we hypothesize that some of these features may identify new photopeaks of interest for burnup analysis. It is also, however, possible that these energies do not correspond to real photopeaks and instead highlight key scattering energies that form the "background" signal from which the ML methods so effectively extract predictive abstract features.

Table 3. Top 24 energies identified by our LIME-based explainability analysis, for the 12-hour cooling condition dataset.

Rank	Energy (keV)	Rank	Energy (keV)	Rank	Energy (keV)
1	891.153	9	36.873	17	280.953
2	446.385	10	55.857	18	893.865
3	443.673	11	896.577	19	205.017
4	337.905	12	66.705	20	899.289
5	118.233	13	61.281	21	272.817
6	278.241	14	318.921	22	888.441
7	180.609	15	58.569	23	316.209
8	286.377	16	340.617	24	69.417

Finally, we performed linear regressions on subsets of the final 24 selected energies to determine the potential for new high-accuracy burnup analysis based solely on classical regression and easily interpretable linear coefficients. Table 4 shows that with as few as 4 features (i.e., the top 4 energies

identified in Table 3), a linear regression model can be fit which predicts with ~80% accuracy the burnup of a fuel pebble, and that this improves to 93% with 8 features.

Table 4. Linear regression on top-N LIME features (all features ~3keV wide).

	↓ RMSE	↓ MAPE (%)	↑ R^2
Top-24 (all)	0.528	5.16	0.9996
Top-8	0.698	6.81	0.9994
Top-4	3.424	20.99	0.9848
Top-2	6.113	58.74	0.9533

We performed the same analysis on our second evaluation dataset, for fuel with 120-hour cooling time and 1-hour measurement time. The top 24 energies selected for that condition are shown in Table 5. We note that although several of the top energies are present for both conditions (12-hour and 120-hour), their order is significantly different, and many additional ones are unique to one condition or the other. This is expected, as the increased cooling time changes the spectral characteristics of the measured fuel, and we have already demonstrated previously that both ML and classical modeling performance is affected by changes in cooling time. Further, we see this as evidence that our ML model very effectively extracts features specific to the dataset and condition on which it was trained, but that these do not naively adapt to external conditions unseen at training time.

Table 5. Top 24 LIME-identified energies for the 120-hour cooling condition dataset.

Rank	Energy (keV)	Rank	Energy (keV)	Rank	Energy (keV)
1	2740.737	9	446.385	17	421.977
2	2754.297	10	443.673	18	66.705
3	888.441	11	2751.585	19	2770.569
4	308.073	12	55.857	20	405.705
5	118.233	13	389.433	21	61.281
6	2735.313	14	348.753	22	351.465
7	337.905	15	584.697	23	69.417
8	58.569	16	354.177	24	454.521

3.3. Additional and future work

Following our LIME-based explainability analysis, we extended our investigation to SHAP-based explainability analysis. This method has thus far proved challenging due to stability issues with training SHAP tree ensembles on our LIME-selected features, but we believe SHAP will provide

useful insight. Whereas LIME is limited to single linear regressions, SHAP supports decision tree models that directly correspond with an explainable data path for a burnup prediction.

Additionally, we have carried out initial work in estimating the uncertainty of our ML models by training ensembles of MLP-based and CNN-based models and estimating their prediction spread. The objective is not to formalize this process to accurately quantify the expected uncertainty of ML-based burnup predictions. However, with this work, as well as a series of candidate burnup thresholds, we expect to be able to directly infer the operational advantage (in fuel usage) of relying on a more accurate ML-based burnup analysis versus a traditional approach.

4. CONCLUSIONS

We have developed a full-core PBR burnup simulation model. Overall, the simulation results from our model agree with the ones from the Origen model built by ORNL. This is an indirect way to validate our simulation model. We expect some experimental measurements will be carried out with actual TRISO particles in FY24 or later so that we can experimentally validate our simulation approach.

We also looked into the simulation workflow in Serpent involving both burnup and photon transportation simulations and was able to setup a procedure to decouple these two simulation processes. This enables us to produce multiple datasets with different collimation and detector configurations, which could be a useful tool for optimizing PBR reactor design.

We were able to accomplish the explainability analysis in the ML models and identify the energy bands that are of importance to the ML model predictions. We also found such energy bands vary with fuel cooling time, which agrees with our early findings in the inference performance of our ML models.

REFERENCES

[1] Akyurek, T., Tucker, L., & Usman, S. (2014, July). Review and charachterisation of the best candidate for isotopes for burnup analysis and monitoring of irradiated fuel. Annals of Nuclear Energy, 69, 278-291.

[2] Butler, K. T. (2018). Machine learning for molecular and materials science. Nature 559.7715, 547-555.

[3] Soto, Carlos X., et al. "A Better Method to Calculate Fuel Burnup in Pebble Bed Reactors Using Machine Learning." *Nuclear Technology* (2023): 1-13.