Advanced Reactor Safeguards & Security

Machine Learning to Improve Burnup Measurement in Pebble Bed Reactors

FY25 Annual Report

Prepared for
US Department of Energy

Yonggang Cui, Carlos Soto, Odera Dim

Brookhaven National Laboratory

November 2025

BNL-5872265

BNL-5872265

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. We also examined the explainability of the MLP regression model. We further improved the burnup model and completed a full-core PBR simulation model in FY2023 so we could validate the simulation with the existing data of well-studied PBR reactors, i.e., PBMR400. In FY2024, we generated a low-variation dataset that still simulated the parameter changes in real reactor operation and verified that the MLP model could give reasonably good estimation of burnup of fuel pebbles. In FY2025, we extended the research to high-variation datasets focusing on variation of cooling time, power density, and transit time. This annual report summarizes the major accomplishments in this fiscal year.

ACKNOWLEDGEMENTS

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

CONTENTS

This page left blank

ACRONYMS AND DEFINITIONS

Abbreviation	Definition
ARSS	Advanced Reactor Safeguards & Security
CNN	Convolutional neural network
DOE	Department of Energy
IAEA	International Atomic Energy Agency
ID	Inventory Difference
LWR	Light-water Reactor
MAPE	Mean Average Percentage Error
MC&A	Material Control and Accountancy
ML	Machine Learning
MLP	Multilayer Perceptron
NN	Neural Network
PCA	Principal Component Analysis
R^2	Coefficient of Determination (square of the Pearson correlation coefficient)
ReLU	Rectified Linear Unit
RMSE	Room Mean Square Error

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 FY2023, we developed a full-core burnup model so that we were able to validate the simulation work against the results from Oak Ridge National Laboratory (ORNL) team and examined the explainability of the MLP model in this application. In FY2024, we improved the burnup model to take into account the parameter variations from the real reactor operation and evaluating the performance of the MLP model in this "operational" scenario. Given the complexity of operational parameter changes and to simplify the problem, we started with datasets with relatively low variances, which also set up the baseline for us to improve the performance of ML models in simulated operational environment. This year, we extended this study to high-variation datasets focusing on potential wide-range changes of cooling time, power density, and transit time in real operation. This report summarizes the results we obtained in this year.In this year, we focus on improving the burnup model to take into account the parameter variations from the real reactor operation and evaluating the performance of the MLP model in this "operational" scenario. Given the complexity of operational parameter changes and to simplify the problem, we started with datasets with relatively low variances, which also set up the baseline for us to improve the performance of ML models in the next phase of this study. This report summarizes the results we obtained in this year.

2. MODELING AND SIMULATION

The effort for this year was focused on trying to get a better understanding of the impact of various simulation parameters to the performance and prediction by the ML algorithms. Although there are a vast range of parameters which ML could be sensitive to, we had selected 3 key parameters based on observed behavior from previous efforts. These parameters are:

1. Transient Time: The transient time refers to the period which the pebbles travel through the core. It is equivalent to time that the pebble experiences burnup over several loops through the reactor core. We varied this by making small perturbations to the time step using randomization algorithms. This also advertently also varies the Effective Full Power Days (EFPD) for the given pebble.
2. Cooling Time: This time refers to the time after each cycle that the pebbles are outside the core under zero power experiencing decay of the various fission product radionuclides. This is typically done to allow some of the short-lived radionuclides to decay away, reducing the noise in the resulting gamma spectra produced by a radiation detector (i.e., HPGe). We varied this by applying a small perturbation around a cooling value.
3. Pebble Power: The pebble power varies spatially in the core due to variations in the local flux as the pebble traverses the core axially. Typically, this is quite different for any pebble which goes through the core as the path they may take is significantly different which leads to different power experienced. We sample a generic power profile of a PBR core and perturbed each power value sampled as the pebble travels through the core by the given percentage.

To investigate the impact of these parameters we varied each parameter independently while leaving the others constant. By doing so we hoped to see how the ML models performed with the variation of individual parameters. We also had a dataset which was regarded as the nominal case where no variation was applied at all. Table 1 below is a description of the configuration of the datasets generated for this effort.

Table 1. Descriptions of 5 datasets isolating variance in 3 key PBR parameters

Dataset name	**Description**
dataset_00	No variations applied
dataset_01A	Up to 20% variation in transient time only (version A)
dataset_01B	Up to 20% variation in transient time only (version B)
dataset_02	Up to 10% variation in cooling time only
dataset_03	Up to 5% variation in power level only

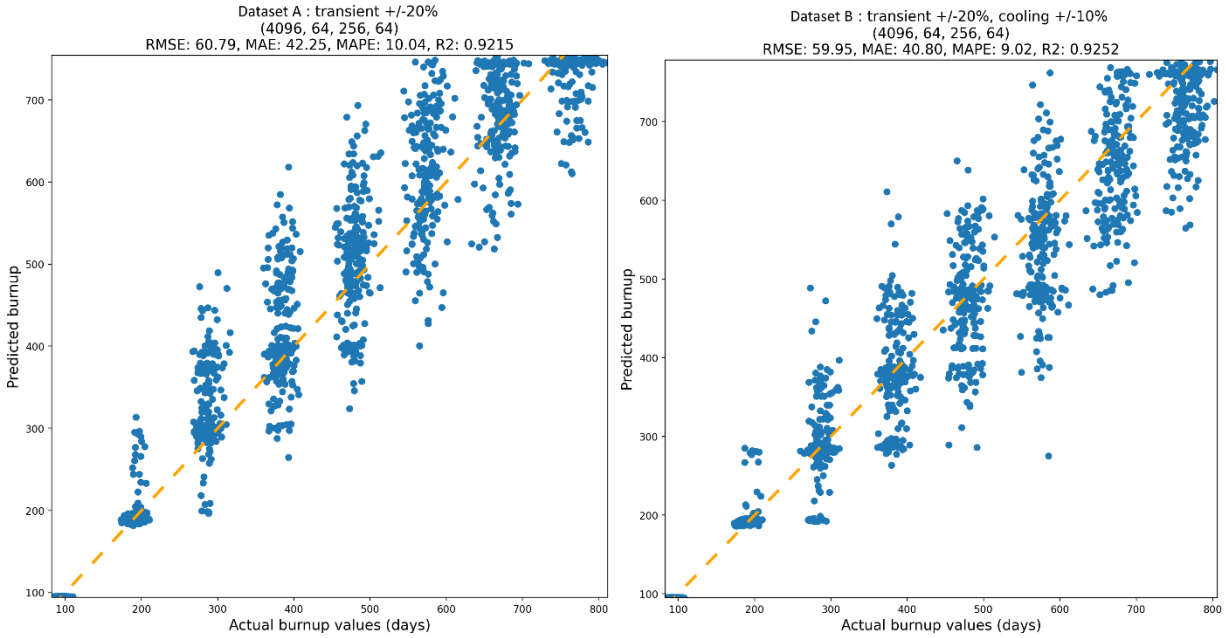
3. EVALUATION OF THE ML MODEL

ML training and evaluation work focused on two core task and two supplemental ones. The core tasks were both systematic determination of ML performance for variations of key PBR parameters (described in Section 2). The supplemental tasks were ensemble-based quantification of modeling uncertainty, and an exploratory evaluation of an alternative model architecture.

3.1. Sensitivity analysis for 3 PBR parameters in combination

As described in Section 2, datasets were created to evaluate impact of combinations of variances applied to three PBR parameters. We directly trained ML models on these datasets to determine the sensitivity of ML-based burnup prediction performance with respect to uncertainty in transient time, cooling time, and power profile. Figure 1 shows the summary of ML results for these three datasets, and Figure 2 shows comparative results vs two photopeak baseline methods (based on Cs-137 activity and ratio of Cs-134/Cs-137 activity).

First, we once again observe that ML prediction performance (accuracy and precision) is significantly stronger than photopeak baselines. Second, we observe two interesting phenomena. First, and quite surprisingly, ML performance appears to improve as additional sources of variation and uncertainty are added (10.04%, 9.02%, and 8.05% error for Datasets A, B, and C, respectively). This is the opposite of the expected effect, as these datasets have increasing and compounding sources of variation. This unexpected finding prompted us to conduct additional experiments reported in Sections 3.2 and 3.3. Second, we observe a slight banding effect in the distribution of ML predictions. We interpret this as the ML model "snapping" to expected enrichment or configurations values for each of the 8 pebble cycles represented during training. We sought to exploit this effect via special ML architecture, described in Section 3.4.



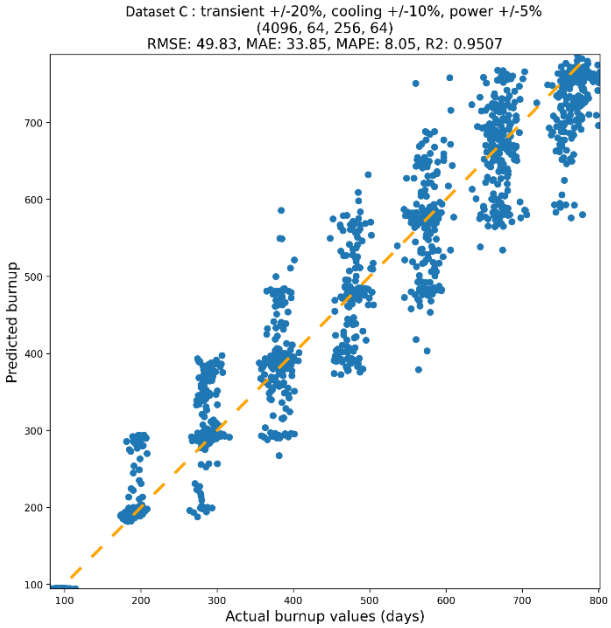


Figure 1. ML performance when trained on Datasets A (*top-left*), B (*top-right*), C (*bottom*).

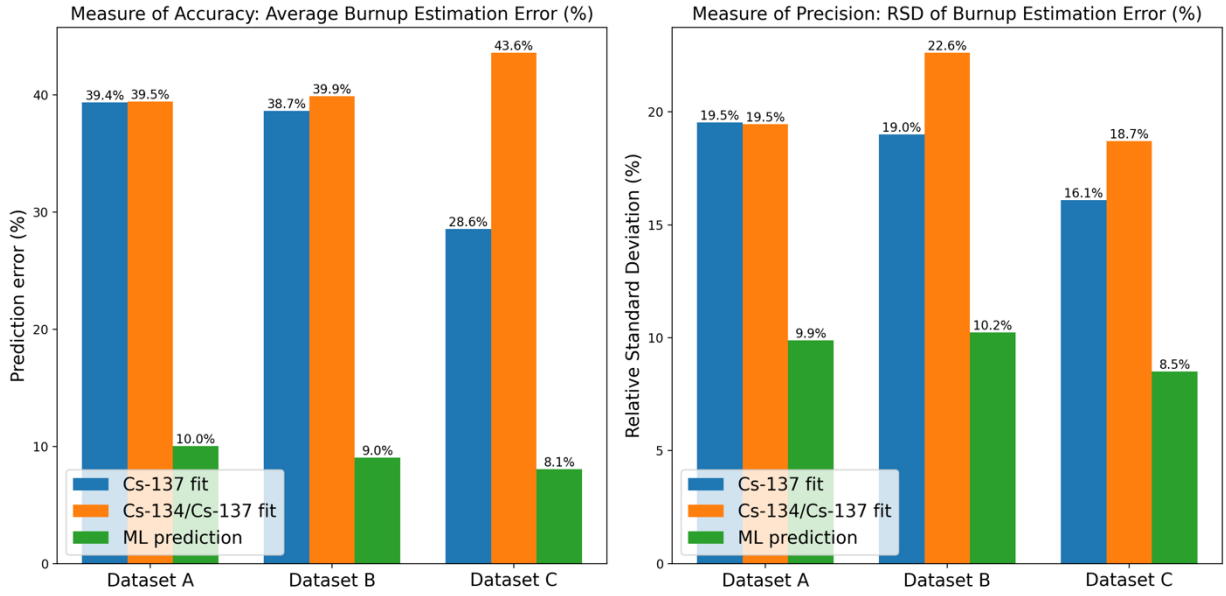


Figure 2. Comparison of ML results vs photopeak methods with two key metrics. The average percent error (*left*, equivalent to MAPE metric in previous figure) is a measure of prediction accuracy. RSD (*right*) is a measure of variance. Lower is better for both metrics.

3.2. Sensitivity analysis for 3 PBR parameters in isolation

Following the analysis described in Section 3.1, we trained and evaluated ML models with the same architecture for new datasets that were produced to isolate the three modeled sources of variance. Table 1 describes the five datasets that were used for this analysis (as well as those in Sections 3.2 and 3.3). The rationale for isolating the sources of variance (rather than iteratively applying additional sources) was due to a desire to identify the relative contribution to ML performance

effects for each of the sources. Note that two versions of dataset_01 were produced, with the same varied parameters.

Figures 2 and 3 show the trained ML results for the baseline and parameter-varied versions of the dataset, respectively. We were again surprised to find that the variations in cooling time and power level (dataset_02 and dataset_03) appeared to slightly *help*, rather than hurt ML burnup-prediction performance (. Although we did not formally explore the causes of this effect, one hypothesis is that variations in these parameters introduce variations in key gamma spectral features that cause the trained ML model to better learn strong representations for those features, or else cause it to be more robust in its support for variance in these features. The applied variations in pebble transient time (dataset_01A and dataset_01B) hurt performance, as expected. Overall, performance differences between the datasets are not large; correlation is strong across the board, and the ML models are able to accurately predict fuel burnup to within 2 to 4 days (MAE, mean absolute error). The primary metric of interest (MAPE, mean absolute percept error) shows consistent prediction errors between 11 and 14%.

Figures 5 and 6 compare ML results with the same photopeak methods (Cs-137 activity and ratio of Cs-134/Cs-137 activity), again showing significantly stronger ML performance, but also showing that dataset_03 yields improvement in photopeak-based burnup estimation as well. This suggests that the additional variation impacts the separability of the Cs-137 and Cs-134 photopeaks too.

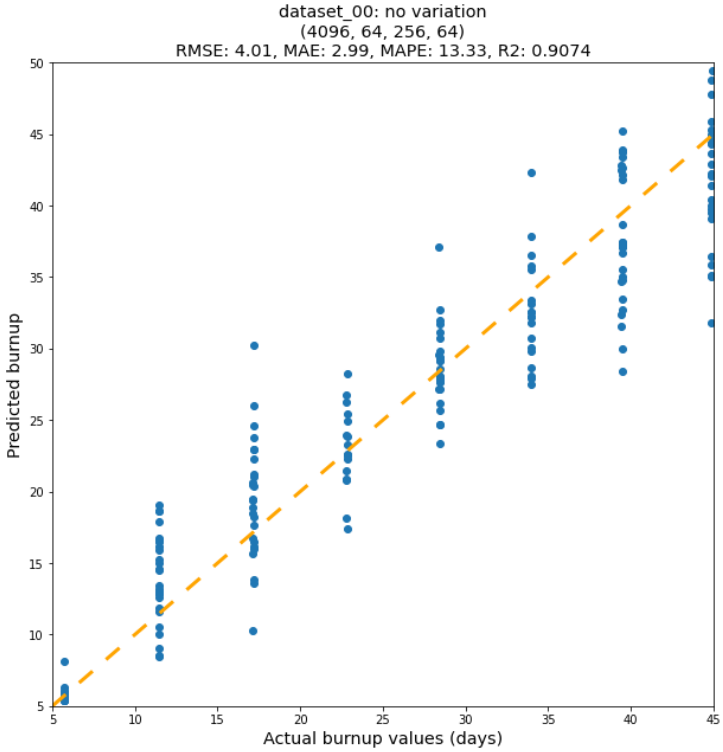


Figure 3. Baseline ML results for dataset with no PBR parameter variations applied

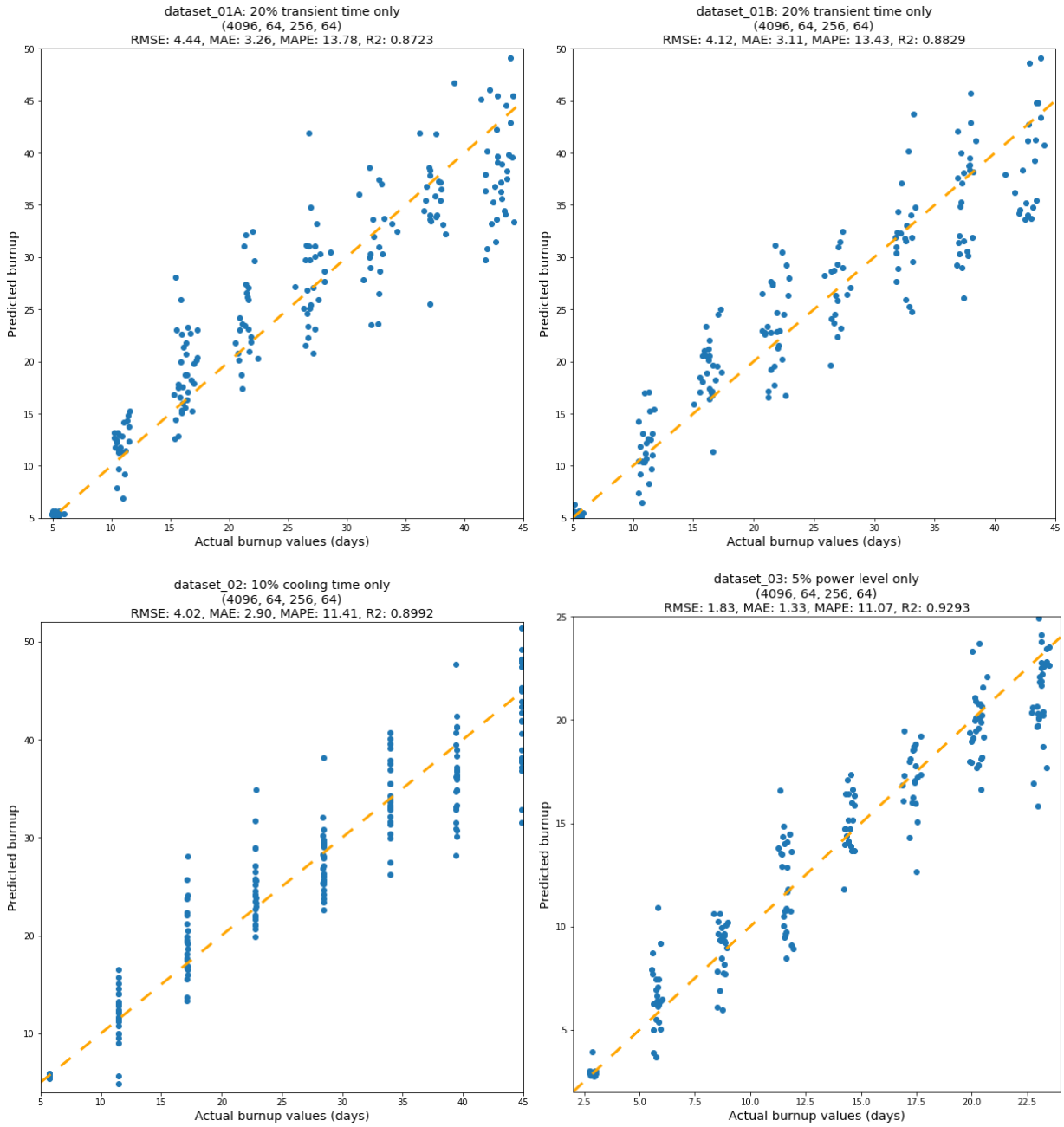


Figure 4. ML burnup prediction performance for isolated variations in 3 key PBR operational parameters: pebble transient time (*top-left* and *top-right*), cooling time (*bottom-left*), and power level (*bottom-right*).

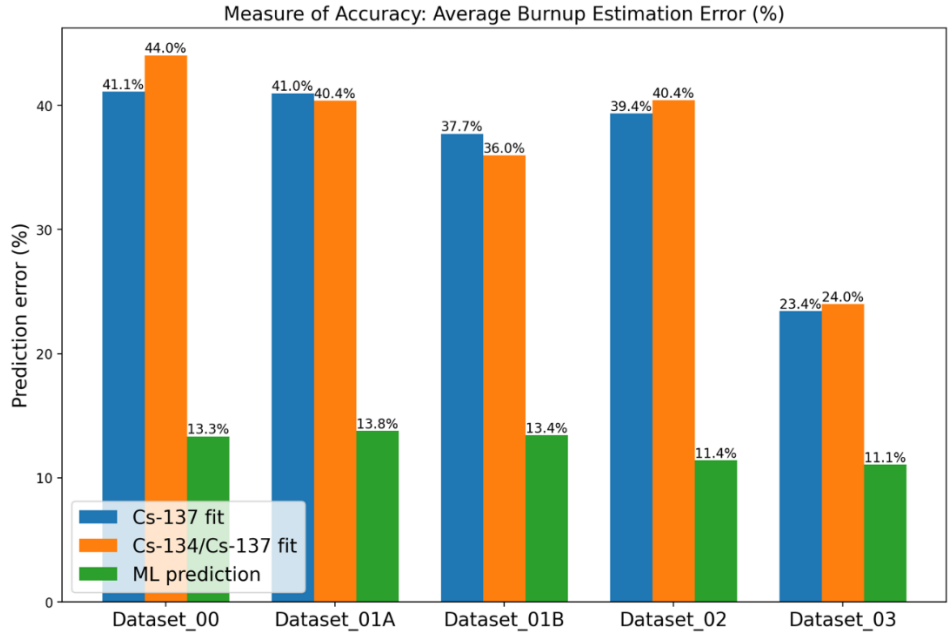


Figure 5. Comparison of ML results vs photopeak methods in average percent error (equivalent to MAPE metric in Figures 3, 4 above) as a measure of prediction accuracy. Lower is better.

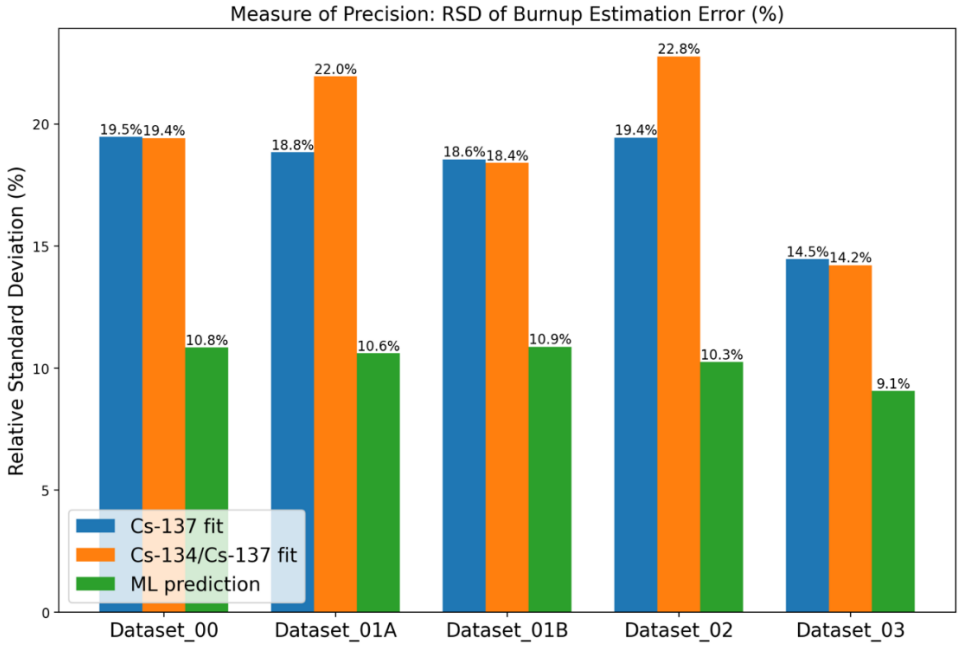


Figure 6. Comparison of ML results vs photopeak methods in Relative Standard Deviation (RSD) of prediction error (a measure of precision and variance). Lower is better.

3.3. Quantifying ML modeling uncertainty

The analyses in Sections 3.1 and 3.2 can be thought of as evaluated uncertainty propagation for the ML modeling effort's source data; in statistics and data science, this is called aleatoric uncertainty. Conversely, the other major source of uncertainty in data modeling efforts is the uncertainty due to the modeling effort itself. This is called epistemic uncertainty, or modeling uncertainty. Various techniques exist to estimate this modeling uncertainty, but many of these require replacing a model design with one that explicitly quantifies and propagates uncertainty as an architectural feature. The disadvantage of these methods (apart from their potential lower modeling capacity and performance) is that they are no longer directly representative of the original ML model used for training and analysis.

The alternative we chose is a simple, but flexible one: model ensembles. Specifically, we use bagging, with random full data subsets, a common architecture and training schedule, and held-out validation sets to track training and convergence anomalies. We used ensembles of size 10 for each of the five datasets evaluated and determined for each the ensemble's mean MAPE (mean absolute percent error) metric, as well as a 95% confidence interval for this value. Table 2 summarizes these findings.

Table 2. ML Ensemble performance and variance for each perturbation dataset

Dataset	**Ensemble's mean MAPE**	**95% confidence interval ($\pm$ 2 σ) in percentage points (p.p.)**
dataset_00	13.23 %	1.55 p.p.
dataset_01A	13.58 %	0.98 p.p.
dataset_01B	13.76 %	1.70 p.p.
dataset_02	12.19 %	2.02 p.p.
dataset_03	10.70 %	1.28 p.p.

The ensemble results confirm the findings in the core ML results: the additional variation or uncertainty in pebble transient time propagates through the ML network and results in very slightly worsened burnup prediction performance, however variation in cooling time and power profile do not worsen ML prediction performance but in fact yield improved accuracy. On the other hand, the ML variance for each of these datasets tracks independently of the gross average performance effect. Burnup prediction variance (represented in Table 2 by 95% C.I.) shrinks for dataset_01A and dataset_03, while increasing for dataset_01B and dataset_02. In all cases, though, there is overlap between the 95% confidence intervals of the perturbation datasets and the baseline. Dataset_03 (variation in power level) comes very close to statistical significance and is a strong candidate for further study. Note: percentage point is used when comparing two percentages as is the case with the data in Table 2 above.

3.4. Exploration of alternative custom ML architecture

The final analysis we performed was an architecture optimization attempt to exploit the banding effect observed in Section 3.1. We designed a modified version of our hybrid CNN-MLP ML model architecture with a classification bottleneck after the CNN backbone and prior to the MLP inference head. The bottleneck would be responsible for attempting to predict the cycle number of the measured pebble – independent of its burnup time – and then make this signal available for the MLP regression head to potentially improve its burnup prediction performance.

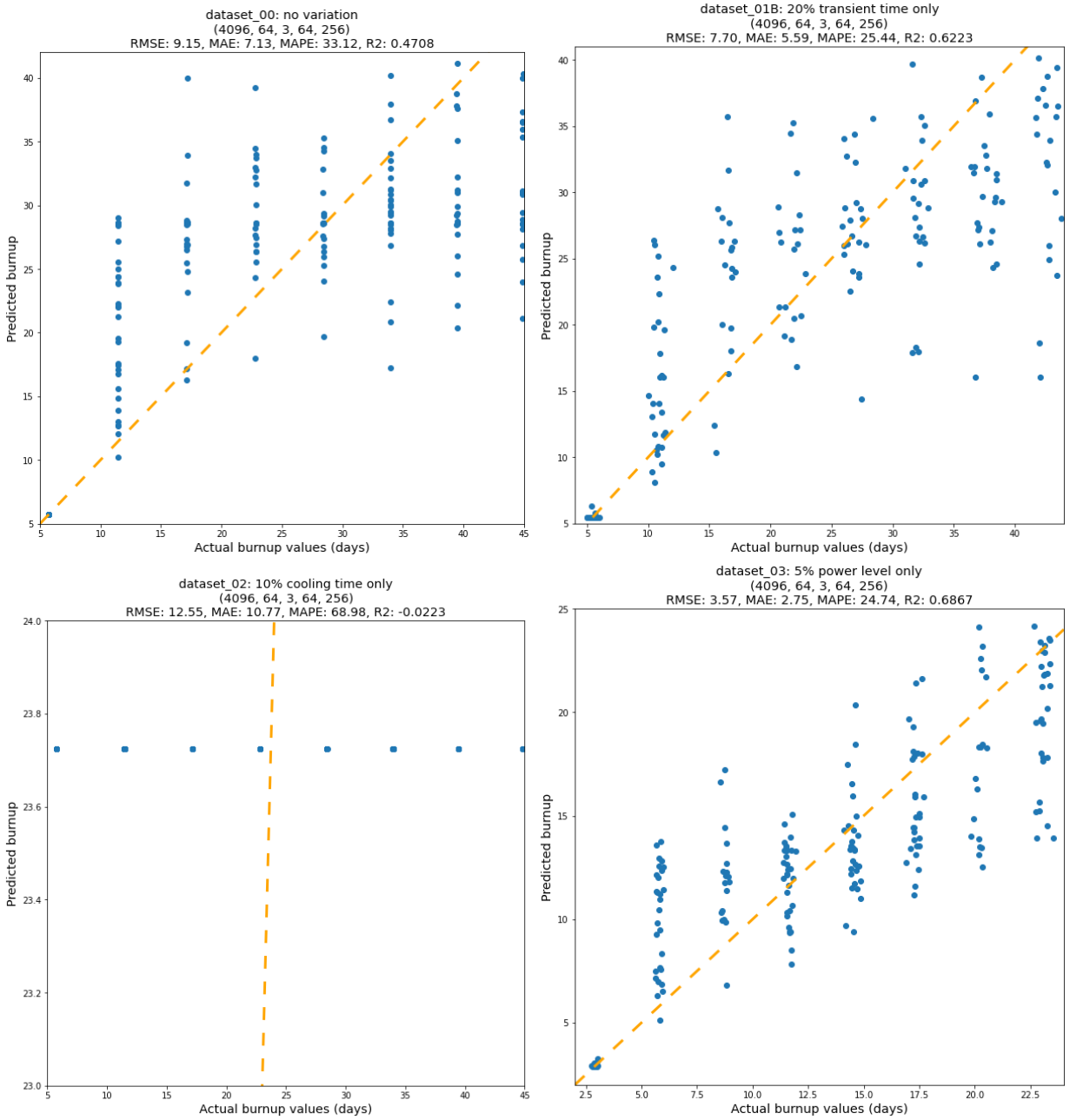


Figure 7. ML prediction results with modified ML architecture, employing a classification bottleneck.

This experiment did not, however, yield the desired results (Figure 7). ML prediction performance degraded significantly relative to the original architecture (Section 3.2), and in one case (dataset_02) the model consistently failed to train. It is possible that further optimization and/or debugging may resolve these issues, but at present it appears that the model architecture we developed is well optimized and difficult to significantly improve upon.

4. CONCLUSIONS

We attempted to investigate how the individual burnup simulation parameters (i.e., cooling time, transient time and power profile) impacted the performance of the ML models. Our ML analysis shows that contrary to expected behavior, with the addition of more variation to the parameters, the ML algorithm seemed to perform better or roughly same. Two main ML analysis were performed.

The first analysis was to investigate the performance of the ML algorithm with cumulative addition of varied parameters (i.e., vary the transient time, then vary the transient time plus cooling time etc.). To better express the performance observed we also compared the results here to the Cs137 and Cs134/Cs137 photopeak ration approach for burnup estimation as shown in Figure 1. The performance comparisons are also expressed as a function of MAPE and RSD. For ML application MAPE is preferred RSD because it accounts for proportional errors relative to ground truth. In any case, regardless of the metric used ML performs better than the Cs137 and Cs134/Cs137 approaches for all cumulative variations investigated, this is shown in Figure 2. We also investigated the performance of the ML algorithm against baseline and obtained a MAPE of about 13.3% which is higher than the 10.04%, 9.02%, and 8.05% MAPE values obtained for the three cumulative variations investigated. This behavior is counterintuitive as we would have expected a poorer performance with additional variations. However, our belief is that ML might be performing better as it views the variations as additional features to latch unto.

The second analysis was to investigate the performance of the ML algorithm independent or individual parameter sets (i.e., vary transient time alone, vary cooling time alone etc.). As with the first analysis, the performance comparison between ML and Cs based approaches are presented in MAPE and RSD metrices. Again, ML performs better than Cs approaches with a lower MAPE of 13.3%, 13.8%, 13.4%, 11.4% and 11.1% respective to each independent dataset listed in order on Table 1. An additional observation is that the transient time variation seemed to be a harder problem for the ML model to predict. This also may indicate that the dynamic random motion of pebbles would significantly impact how well ML would perform.

In future, we would like to use burnup in (MWD/kg) as against time in (days). The nuance being that time in days is linear while MWD/kg is nonlinear due to the local fluxes and power being different at different steps for a given cycle. In addition to this the heating of a specific pebble varies as the fission products build up and concentrations change for that specific pebble. This might also account for some of the banding effects observed in the ML results.

REFERENCES

[1] Akyurek, T., Tucker, L., & Usman, S. (2014, July). Review and characterization 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.