# Advanced Reactor Safeguards & Security

# *Use Machine Learning to Improve Burnup Measurement in Pebble Bed Reactors*

## *FY24 Annual Report*

**Prepared for**
**US Department of Energy**

**Yonggang Cui, Carlos Soto, Odera Dim**

**Brookhaven National Laboratory**

**September 2024**

**BNL-NN-20240930-00411-00-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. 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 are extending the burnup model to take into account possible parameter variations from real PBR reactor operation, which includes generation of a new dataset and evaluation of the ML model performance on the new dataset. 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 annual report summarizes what we have accomplished in these two tasks by 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. Last year, 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 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

For proof of concept and saving computing time, our inception preliminary burnup models were simplified. For example, the cooling time was set to a fixed value that cannot be controlled precisely in real reactor operation. To generate more spectra within a limited time, we sampled the compositional information of the pebbles after the burnup at each individual zone, which is not the

case in real operation as the pebbles can only be measured when they are ejected from the core. In addition, our early models also assumed lower variations of residence time and power of the reactor core.

After talking to several industrial reactor designers, we decided to expand the burnup model to consider more and larger variations in operational parameters to simulate real operational scenarios. Figure 1 shows the sketch of the new model. A few changes were made in this model compared to the previous ones:

- We considered a sine power profile with a peak power density of 0.057kW/g. We divided the core into five burnup zones. In this case, the top and the bottom zones have close to zero power. Especially, the bottom low power zone matches the real PBR reactor design.
- We allowed each fuel pebble to pass through the core 8 times with an averaged residence time of 100 days. In each zone, we allowed the residence time to vary ±5%, resulting in a simulated residence time of 100±5 days.
- Then, we also varied the cooling time by ±5%, assuming that the pebble sorting system can control the time precisely. The cooling time was set to 24 hrs resulting with a cooling time of 24±1.2hrs. Such cooling time would mostly affect the very short half-life isotopes concentration. We also moved from 12 hr to 24 hr cooling times because Serpent's precision on the burnup days is to one decimal place. This means that cooling times less than a day may in some cases be rounded down resulting in zero cooling time. This also occurs especially because we use small variations like ±5% on the cooling time.

Using this model, we were able to produce two datasets of 800 gamma spectra and one larger dataset of 4000 spectra of burned pebbles. We were able to optimize the script for automating the pebble emission source generation allowing us to produce the data each dataset in 3 days. This will further be improved with the acquisition of new cluster machines for performing simulations.

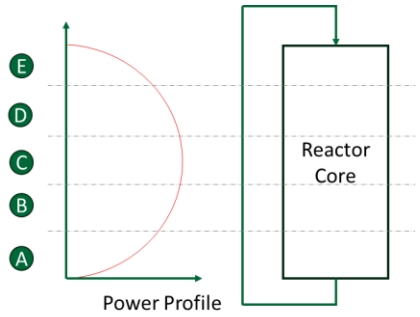


Figure 1. New burnup model with a sine power profile and five burnup zones

## 3. EVALUATION OF THE ML MODEL

We evaluated the multi-layer perceptron (MLP) and convolutional (CNN) neutral networks that we developed in the first year of this project on the new, more operationally relevant data containing 800 spectra of pebbles at various burnup steps. We split the dataset into subsets for ML training, validation, and testing with ratios of 80:10:10, and began by retraining the previous model with similar hyperparameters and training routines as those established with previous datasets. Following this, we moved on to further optimization and comparative studies.

### 3.1. Overcoming convergence issues

We initially encountered convergence issues with model training, which persisted for various hyperparameter changes and optimizations. After further investigation, we found that training on the new dataset was particularly sensitive to the activation functions used in the neural network model. We determined that the ReLU (rectified linear unit) function led to the most stable training for a range of MLP model depths (from 0 to 4 hidden layers) and widths (8 to 256 neurons). This was slightly surprising, as ReLU exhibits only weak nonlinearity and is most often useful for very deep NN models, such as those used for image and language processing. Figure 2 shows examples of the poor model convergence behavior observed prior to the switch to the ReLU activations.

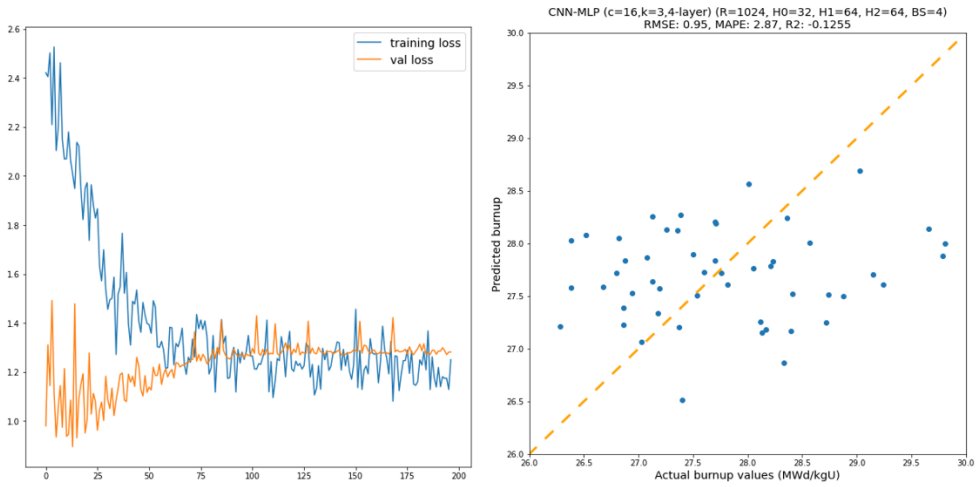


Figure 2. Left: Training and validation loss for CNN-MLP hybrid model that failed to converge, prior to changes of the NN activation function. Right: Resulting uncorrelated model predictions demonstrating regression to mean phenomenon.

In parallel with this NN convergence resolution effort, we carried out two additional investigations looking at the impact of feature selection and feature density on classical regression performance, and the distribution of data variance at different spectra energies. The motivation for both directions was to better understand the effects of increased data variance in the new dataset on model predictive performance. For the first part, we applied principal component analysis (PCA) to construct orthogonal proxy features as linear combinations of the original spectra energy counts, ranked by contributed variance. This enabled us to estimate the number of unique constructed features ("principal components") that account for the majority of the dataset variance. We found that a single principal component accounts for greater than 82% of the dataset variance and 11 components account for 97%; the top 64 components account for 99.37% (see Figure 3). In the following subsections, we demonstrate linear regression performance with these idealized features. For the second part, we constructed a spectral dampening filter to explicitly reduce the impact of data at high-energy end known to contribute high variance but comparatively little information. The construction of this filter (Figure 4) was motivated by the approximate energy ranges in gamma spectra that are known or expected to yield reliable photon counts or correspond with non-noise measured true signal. We used this filter to investigate the impact of reducing this anticipated source of uninformative variance (i.e. likely noise).

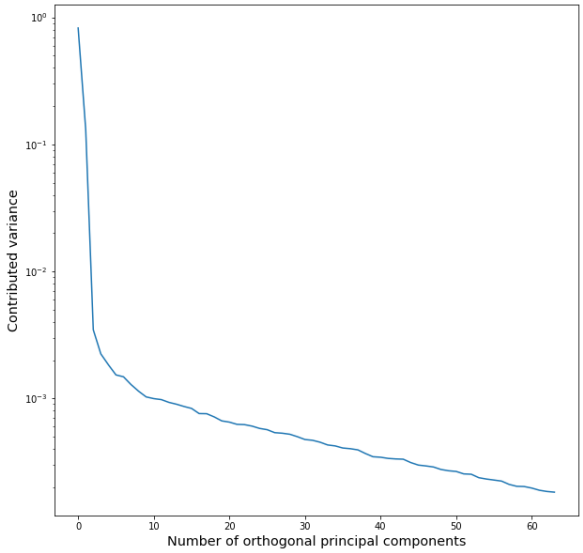


Figure 3. Contributed variance for top 64 constructed principal components, used to investigate linear regression performance with limited variance; note the log scale on the vertical axis.

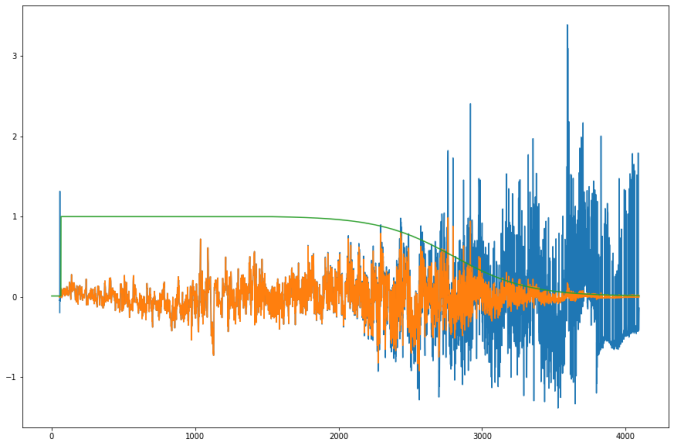

Figure 4. Spectra dampening filter used for exploration of variance contributions to ML model training at different energies.

## 3.2. ML and baseline performance on full dataset

Following the identification of the dataset sensitivity to activation function selection and the choice of the ReLU (3.1) activation function, we re-performed a series of hyperparameter optimizations to determine strong architecture and hyperparameter choices well suited to ML-based prediction of burnup with the new dataset. In particular, we revisited the rebinning strategy employed as part of our data pre-processing prior to ML training and prediction. Whereas we previously observed highly robust training with a range of rebinning ratios, we found with the new dataset that performance was markedly stronger with small rebinning factors. Figure 4 demonstrates the accuracy achieved with the optimal model shape we observed in our testing (4-layers, widths 64, 32, 16, 8). The model achieved a root mean square error (RMSE) of 1.57, and a mean absolute percent error (MAPE) of 12%. This

performance is poorer than what was achieved on the earlier, simpler datasets. However, this was expected as the spectra were produced via more representative core modeling than the prior lattice simulations, representing more complex features and behavior.

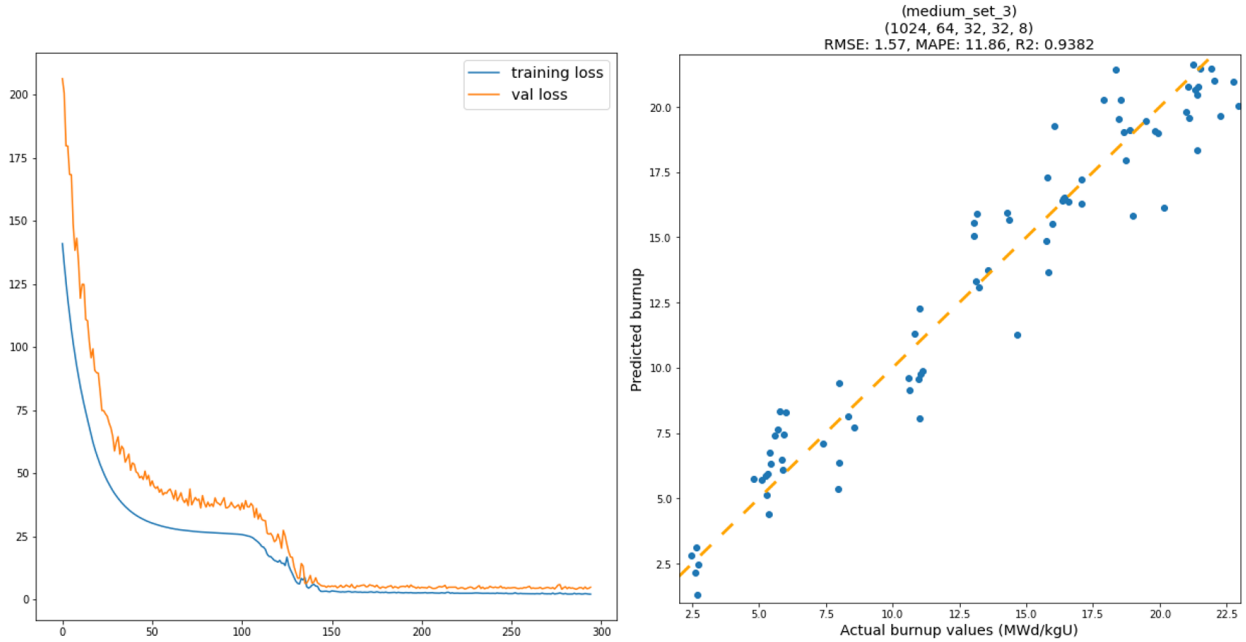


Figure 5. 4-layer MLP model trained on new dataset, with 4x rebinning, and all ReLU activation functions. Left: training for 300 epochs, right: test prediction accuracy.

Relative to the ML models that proved successful with prior datasets, we found that deeper models with lower rebinning both converged better during training and performed better in tests with the new spectra. We noted that while the ML model performance may still be under-optimized (as working to resolve convergence issues limited our optimization efforts), it still performs strongly, particularly compared to the series of baselines we prepared. We compared the results with three separate baselines, augmented from our prior baselines in two ways. First, we applied the noise dampening filter we designed to smoothly discount the high-energy high-variance portion of the spectra (as well as clamp out the very low energy noise, below approximately 40 keV). We found in initial testing that this filter significantly improves the performance on the baseline linear regression methods with the new high-variance data. On the other hand, we found that this filter has no effect on ML model performance. This can be explained in two ways: the preprocessing for the ML model includes a normalization step which largely negates the filter's effect, and the ML model implicitly learns which features are informative and correlated with the prediction target and so benefits less from a heuristic filter choice. The second augmentation relative to our earlier baselines is the inclusion of PCA-based constructed features as a companion test for the effects of high variance on linear regression performance – as well as the establishing of a repeatable linear baseline that is independent of manual feature selections.

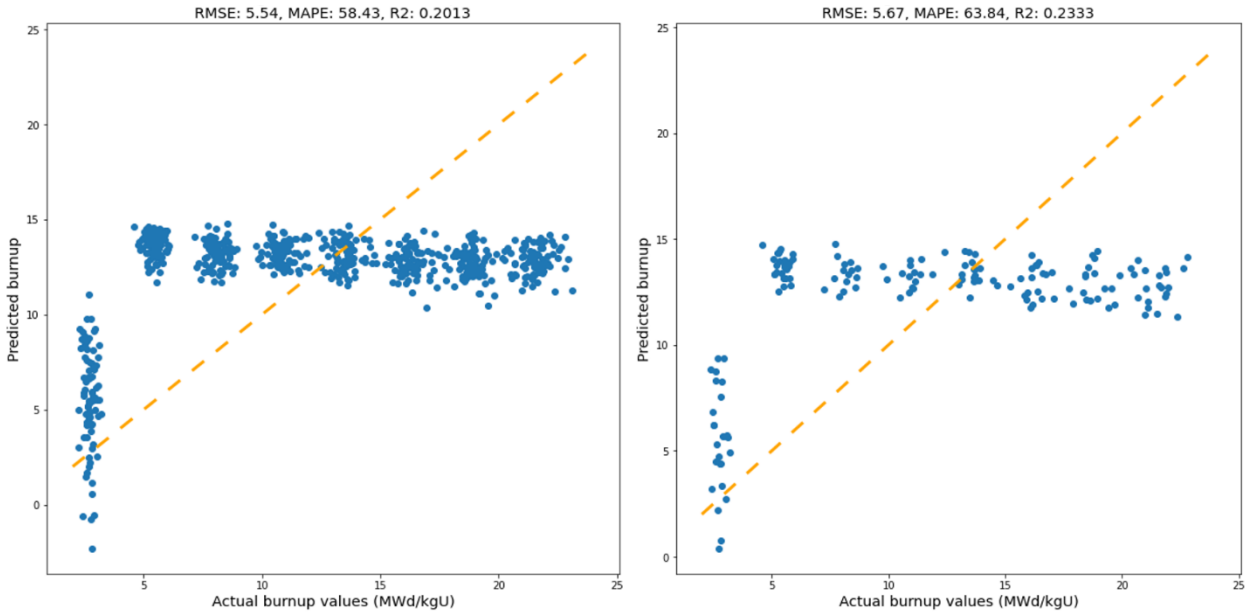


Figure 6. Linear regression baseline with Cs-137 photopeak at 662keV. Training subset left, test subset right.

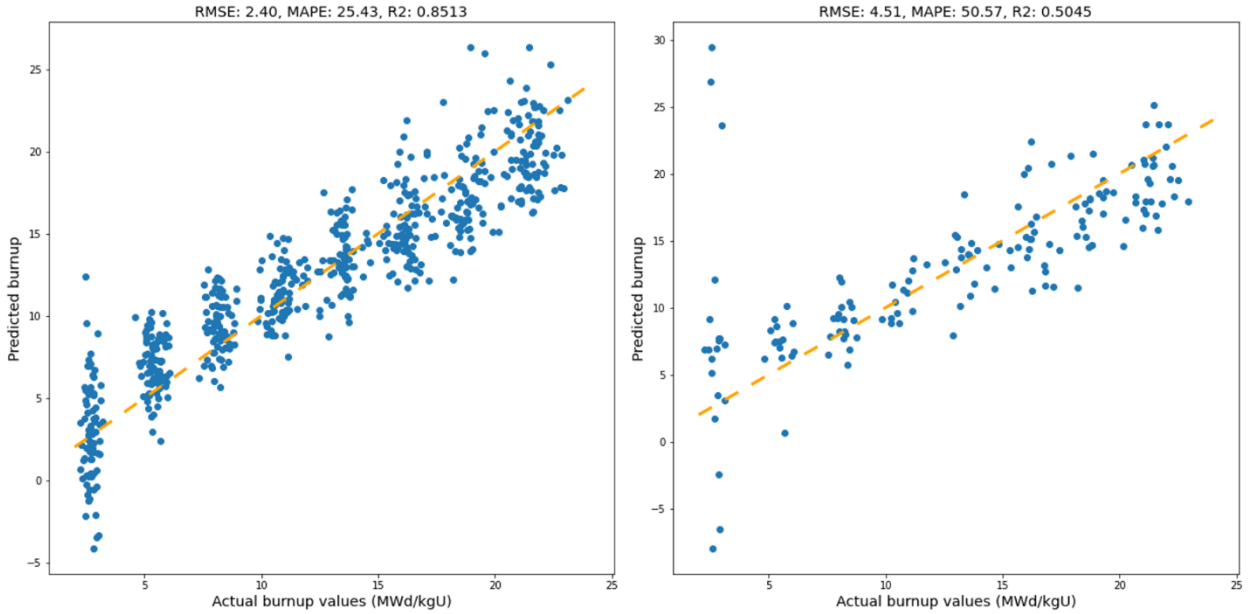


Figure 7. Linear regression baseline with 78 established photopeaks of potential relevance. Training subset left, test subset right.

Figures 6 and 7 show the results of the standard linear regression baselines with single (662 keV) and multiple (78 peaks identified in [Akyurek 2014]) photopeaks. Apart from the poor fit and prediction performance (51 – 64% error), the most noteworthy aspect of these results is the obvious outlier effect apparent in low-burnup spectra samples. This is most apparent for the single peak regression – in which case it completely disrupts the ability to fit the data – but it also has a visible effect on the multi-peak regression. While the fit in that case is significantly better for the training subset than it was for

the single-peak regression, the impact on fit quality and generalization is still very visible in the test subset, where low-burnup samples have wildly varying predictions.

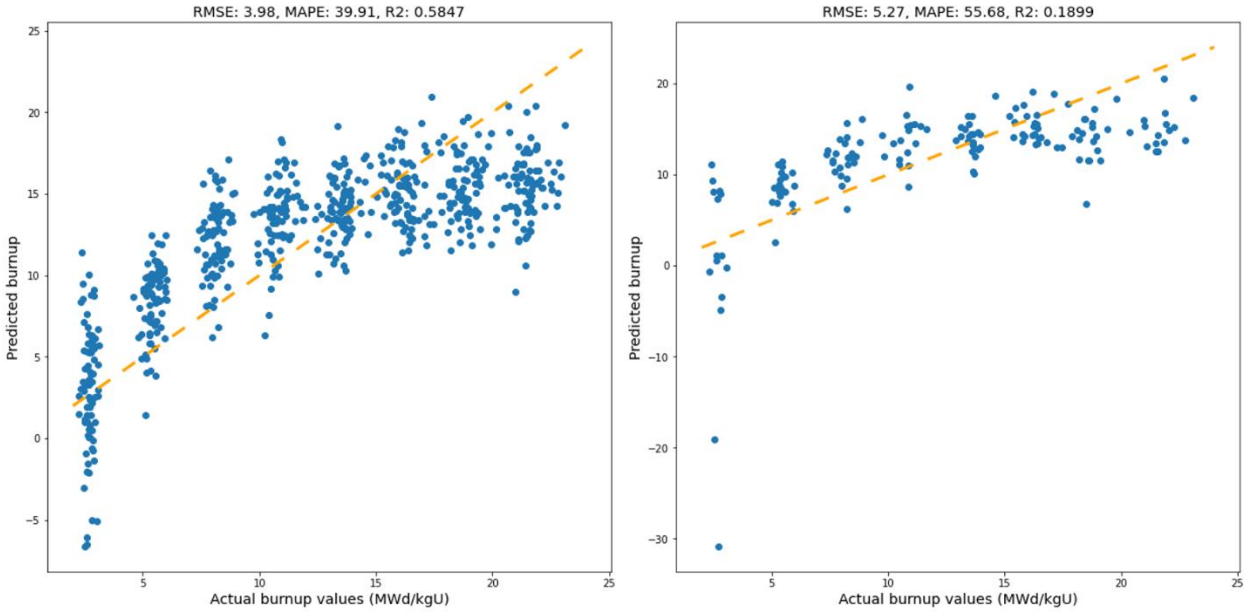


Figure 8. Linear regression with PCA-constructed proxy features (using 64 principal components). Training subset left, test subset right.

This low-burnup outlier effect is further validated by the linear regression performance on PCA-constructed proxy features. This test transforms the raw spectral data into an intermediate representation in which each constructed feature is a linear combination of all the spectra features, but weighted such that all features are fully orthogonal (i.e. uncorrelated) to one another. This approach enables us to select a subset of features in a domain-independent manner while emphasizing feature contribution to dataset variance. The intent is not necessarily to use these PCA-features in place of known isotope photopeaks, but rather to establish an independent baseline that ought to be representative of linear regression performance with strong feature selection criteria. We found in this case that linear regression with the PCA features (using the top 64 principal components) exhibits the same pronounced outlier effect for low-burnup samples as did the single- and multi-peak regression baselines. The impact on prediction accuracy for low-burnup samples is smaller, but the regression fit is skewed more than for the multi-peak regression. Ultimately, the PCA-feature regression performs slightly poorer than the multi-peak regression (56% error).

## 3.3. ML and baseline performance on timestep-filtered dataset

Due to the pronounced impact on baseline regression methods' prediction performance of the low-burnup outliers, we filtered the dataset to remove these and reperformed our analyses. The first timestep in the core modeling was removed; this accounted for pebble samples with up to approximately 110 days of burnup (approximately 3.1 MWd/kgU). The reason for the sample disparity for this first timestep is]since in the very first step of the burnup of a given pebble the used fuel isotopics (i.e., fission products) are very small for that short step in comparison to other subsequent step (this effect is cumulative following the fission and decay chain of the isotopes in our burnup vector), hence when we source scale in GADRAS to avoid dead time effects, this first step amounts

to a small activity and spectral bins with high uncertainty. Note that we have always kept the GADRAS source scale applied consistent from the inception of this project. With this subset removed, we re-fit and re-trained the baselines and ML models on the remaining 700 spectra.

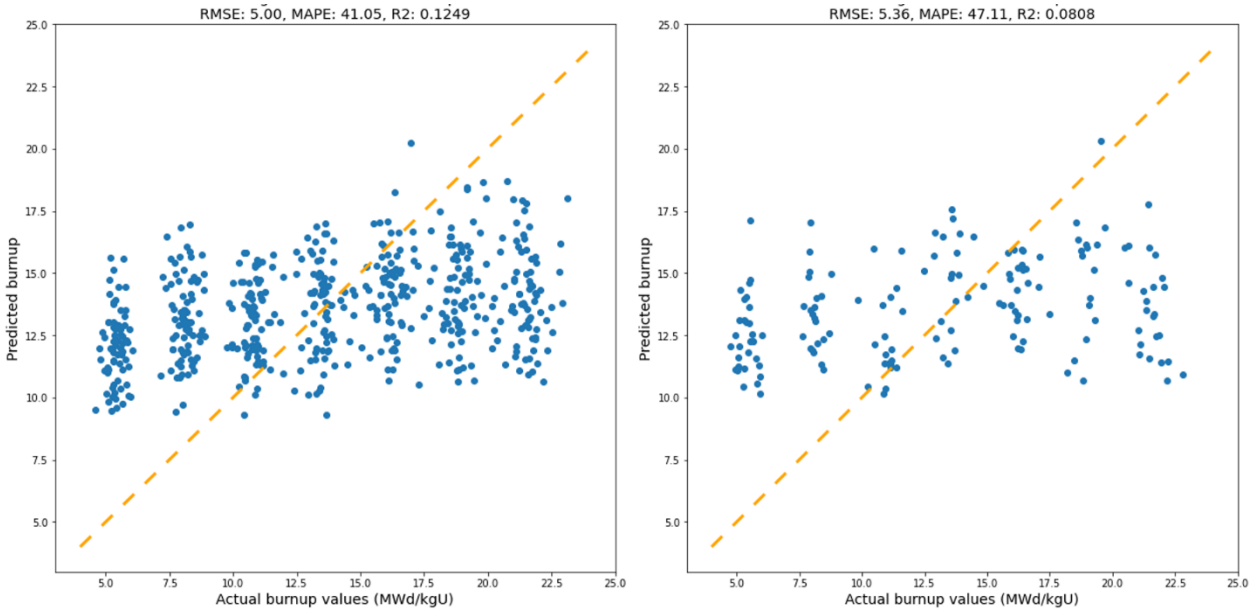


Figure 9. Linear regression baseline with 662 keV peak, first time-step removed. Training subset left, test subset right.

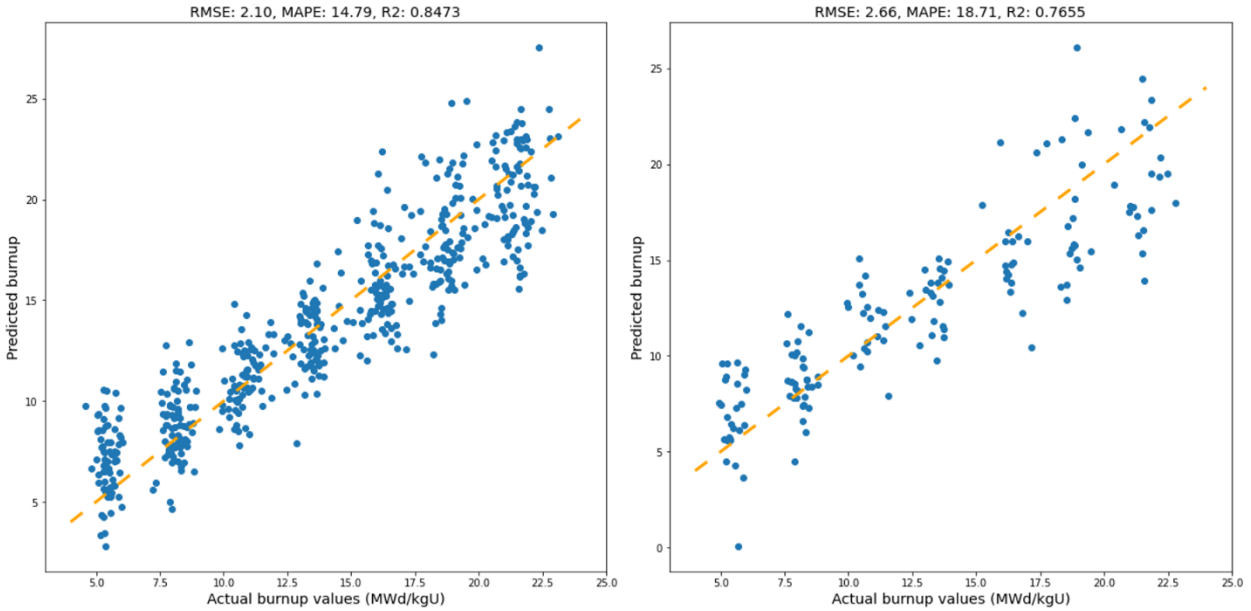


Figure 10. Linear regression baseline with 78 photopeaks, first timestep removed. Training subset left, test subset right.

With the first timestep removed, the observed outlier effect was correctly resolved, and we can see the true baseline performance (Figures 9 – 11). Single-peak regression quality is still very poor (47% error), but the multi-peak regression improved significantly, achieving prediction errors as low as 19%. The PCA-feature regression predictive performance was very similar (18%) to the multi-peak

regression, indicating that this set of selected photopeaks is a strong set of feature candidates for linear regression, and can perform comparably to constructed features that are forced to be uncorrelated and mutually informative.

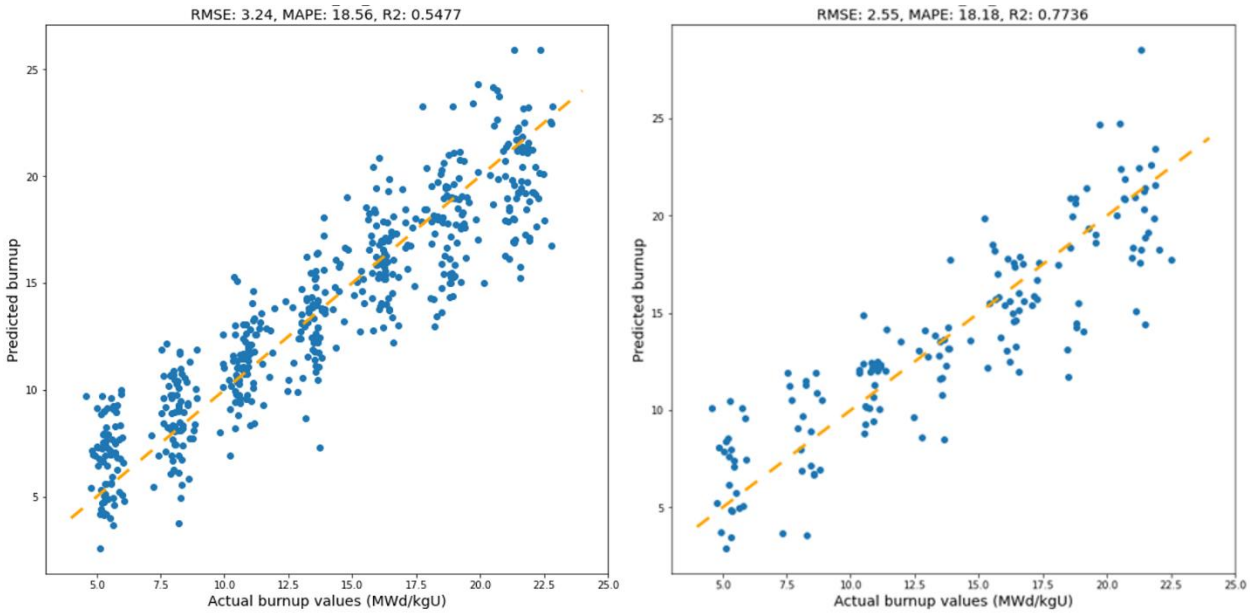


Figure 11. Linear regression with PCA-constructed proxy features, first timestep removed. Training subset left, test subset right.

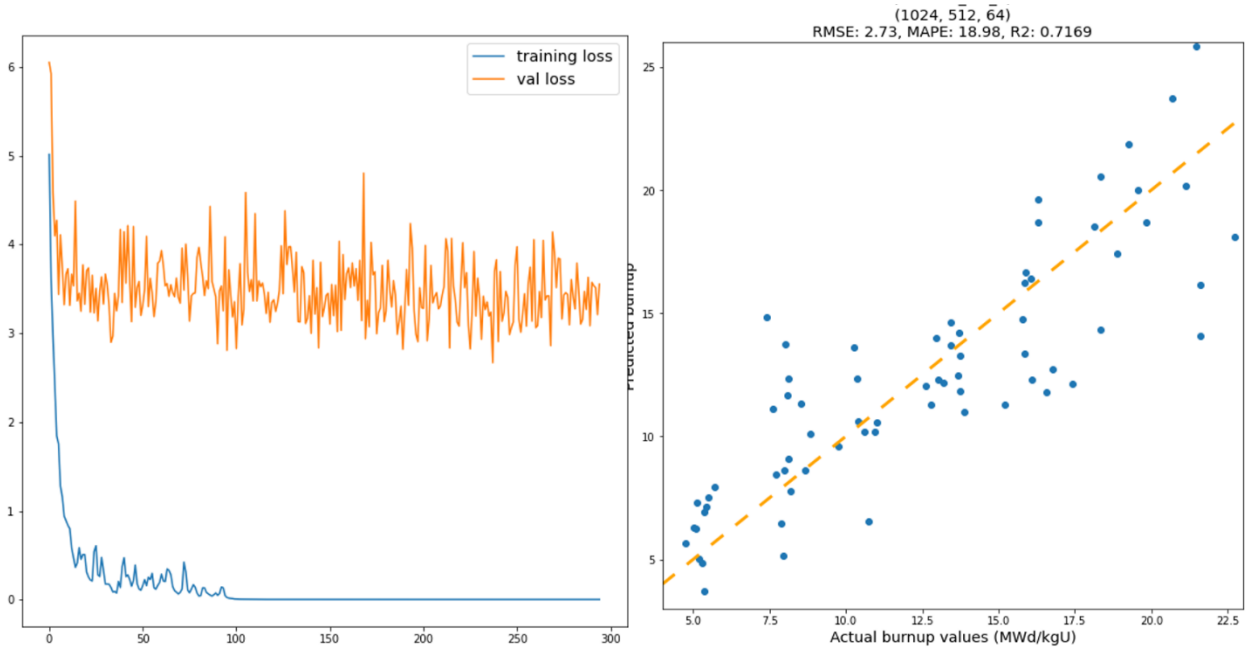


Figure 12. 2-layer MLP model trained on new dataset, with first timestep removed, with 4x rebinning, all ReLU activation functions. Left: training for 300 epochs, right: test prediction accuracy

However, in contrast to the three linear baselines, the ML model re-trained on the low-burnup-filtered subset resulted in poorer performance than on the full dataset (Figure 12, 19% error vs 12% for full

dataset). In fact, it performed comparably to the multi-peak and PCA-feature regressions. Re-examining the results for the full dataset (Figure 4), the ML model performs very well on the problematic first-timestep spectra samples (up to ~110 days of burnup) that proved very challenging for the linear baselines, but which raised the overall accuracy of the ML model.

Our interpretation of these results is that the ML model remains under-optimized for the new data, and that headroom remains in the accuracy it can achieve. Due to challenges with model convergence and then with outliers affecting the effectiveness and fair comparison of the baselines, our optimization efforts for the ML models were limited. We note, for instance, that the training characteristics for the ML model on the low-burnup-filtered subset (Figure 11, left) suggest that the model is again not fully converging and that changes to the ML models and training routines (hyperparameters, learning rate schedules, optimizers, hybrid models) may overcome the current predictive performance plateau.

## 4. CONCLUSIONS

We have extended our burnup model to consider the variation of operation parameters in a real PBR reactor. The model allows us to produce simulation data close to operational conditions. As a start, we used this model to produce a relatively low variance dataset this year containing 800 spectra. We were able to overcome the convergence issues in the ML model training. The trained model performed comparably to the baseline methods if the low burnup outliner data points were removed from the dataset. However, the ML method was more robust and performed better with the entire dataset. We concluded that the ML model remains under-optimized for this new dataset. In FY25, we will continue the optimization work as well as test the ML model with high variance datasets.

## 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.