**Statistical Determination of Spent Nuclear Fuel Origin**

Brian O’Neil, Anthony Scopatz

*University of Texas at Austin, 1 University Place R9925, Austin, TX, 78752, corresponding author: brian.oneil@mail.utexas.edu*

# INTRODUCTION

The ability to determine the origin used reactor fuel has important applications in nuclear forensics. Because of its high radiotoxicity, used fuel (UF) poses a threat to human health, and makes an ideal component of a weapon designed to disperse radiotoxic material. In the event that used fuel is seized during illicit transport, or if a device containing used fuel were discovered or detonated, it is of immediate importance to determine the origin of the fuel so that appropriate nuclear security measures can be put in place.

Here, the isotopic compositions of spent fuel from 118 historical U.S. power reactors are used to train two different statistical classification algorithms. These algorithms each attempt to determine the origin of a used fuel sample.

A set of training samples was generated by using data obtained from the U.S. Energy Information Administration (EIA). The EIA data includes the fuel burn histories for each core. These data were subjected to a burnup simulation methodology [2].

The origin determination is posed as a statistical classification problem over 118 classes. Each sample consists of a feature vector containing the mass fraction of each of the 157 isotopes tracked in the burnup simulation.

The goal of this work is to provide a robust and accurate method to determine the origin of a used fuel sample by its isotopic composition.

# DESCRIPTION OF THE ACTUAL WORK

**Training and Test Data Generation**

One problem that arises when evaluating techniques to identify spent fuel is the relative difficulty in obtaining training and test data. Many analytic tests exist to determine spent fuel composition including mass spectrometry, gamma spectrometry, and alpha spectrometry. However, because of the high radiotoxicity as well as security considerations, it is difficult to obtain and test actual specimens. Therefore, it is common to use validated simulation techniques to determine the composition of hypothetical used fuel. Traditional neutron transport and depletion simulation techniques are computationally expensive, which limits the number of simulations that are practical to use in classification.

A consideration that differentiates the current work from previous work is the size of the available data set for training and characterization. This is enabled by the technique introduced in [2] which was developed to rapidly compute isotopic inventories for nuclear fuel cycle applications. This method made it computationally tractable to use the burnup profiles for all fuel burned in all U.S. commercial power reactors to produce simulated isotopic vectors for every fuel assembly used in a U.S. reactor. Thus the resultant database provides the isotopic compositions of tens of thousands of fuel assemblies that can be used to train statistical classifiers.

**Methods**

The problem of classification with a large number of training samples can be divided into two parts:

* Density Estimation
* Classification

Two different methods are presented in this work. Each takes a different approach to both parts of the problem. The first method uses Maximum Likelihood Estimation (MLE) for density estimation and a Bayesian classifier to perform classification. The second method is a principal component analysis (PCA) nearest-neighbor (kNN) algorithm that combines the tasks of density estimation and classification into a single algorithm.

# RESULTS

# MLE/Bayes Classifier

The MLE/Bayes method used 50 randomly selected training samples (fuel assemblies) from each reactor in the database to estimate the isotopic composition distributions. 100 “unknown” samples were randomly selected for classification.

With the raw data of the selected samples, MLE was not helpful because the resulting covariance matrices were singular to computational precision. To remedy this, the data was filtered to include only isotopes with mass fractions greater than 10-4. This resulted in non-singular covariance matrices, allowing the Bayes classification to proceed.

The error rate was 0.9915, precisely the same error rate that would be expected from randomly guessing at the classification. This suggests that among the isotopes remaining after the filter, there was simply not enough variation between simulated reactors to make a valid classification using this methodology.

# k-Nearest Neighbor

The PCA/kNN classifier also used the same 50 training samples as the Bayes classifier for the kNN prototypes. It was used to classify the same 11800 fuel samples, 100 for each reactor. The number of eigenvectors used to define the classification eigenspace was varied between 10 and 100. The number of nearest neighbors in the kNN algorithm was varied between 1 and 100.

The resulting error rate was 0.9890, and thus was not significantly better than random sampling. It is very unlikely that this method would perform appreciably better than the minimum-error-rate classifier used in the Bayes algorithm unless the isotopic distributions were distinct enough that the non-parametric density estimation technique captured the statistics better than the assumed normal distributions of the MLE/Bayes classifier.

**Summary**

Multivariate statistical analysis of the kind presented here cannot identify a specific reactor facility based upon the isotopic composition of a sample of used fuel. The primary difficulty in the work presented here is that the differences in the physical processes that affect the isotopic composition vary only in the most subtle ways between the reactors being classified. This is in stark contrast with previous work that can only classify spent fuel into very broad categories of reactor type.

**Future Work**

The problem of origin determination of spent fuel is a pattern recognition problem of enough importance to warrant future study. The techniques described here could not identify a specific reactor facility. There are too many classes and not enough variation in the isotopic compositions. However, these techniques may be able to extend some previous work, starting with the ability to differentiate between a BWR and a PWR. Additionally, it may be possible to use such methods in conjunction with previous work to identify a specific reactor facility after certain features of the facility have been determined by other methods. The methods of [1] and [3] can accurately determine initial fuel composition, reactor type, cooling time, and burnup. These features could then be analyzed to identify a specific reactor, or at least a restricted set of them.

# REFERENCES

1. G. Nicolaou, “Identification of unknown irradiated nuclear fuel through its fission product content”. *Journal of Radioanalytical and Nuclear Chemistry*, 279(2):503-508. 2009.

2. AM Scopatz, EA Schneider. “[A new method for rapid computation of transient fuel cycle material balances](http://www.mendeley.com/c/2765215882/p/1362681/scopatz-2009-a-new-method-for-rapid-computation-of-transient-fuel-cycle-material-balances/)”. Nuclear Engineering and Design. 239:2169-2184. 2009.

3. M. Scott, Nuclear Forensics: “Attributing the Source of Spent Fuel Used in an RDD Event” Master’s Thesis. Texas A&M University. 2005.