Todo list

Evaluating Statistical Methods for Nuclear Forensics Analysis

by

Arrielle C. Opotowsky

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Engineering Physics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2017

Date of final oral examination: 01/01/2100

The dissertation is approved by the following members of the Final Oral Committee:

Jane Doeverything, Professor, Electrical Engineering

John Dosomethings, Associate Professor, Electrical Engineering

© Copyright by Arrielle C. Opotowsky 2017

All Rights Reserved

Please insert your dedication here.

ACKNOWLEDGMENTS

It is customary for authors of academic books to include in their prefaces statements such as this: "I am indebted to ... for their invaluable help; however, any errors which remain are my sole responsibility." Occasionally an author will go further. Rather than say that if there are any mistakes then he is responsible for them, he will say that there will inevitably be some mistakes and he is responsible for them....

Although the shouldering of all responsibility is usually a social ritual, the admission that errors exist is not — it is often a sincere avowal of belief. But this appears to present a living and everyday example of a situation which philosophers have commonly dismissed as absurd; that it is sometimes rational to hold logically incompatible beliefs.

— David C. Makinson (1965)

Above is the famous "preface paradox," which illustrates how to use the `wbepi` environment for epigraphs at the beginning of chapters. You probably also want to thank the Academy.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

FIXME: basically a placeholder; do not believe

I did some research, read a bunch of papers, published a couple myself, (pick one):

1. ran some experiments and made some graphs,

2. proved some theorems

and now I have a job. I've assembled this document in the last couple of months so you will let me leave. Thanks!

1 INTRODUCTION

The realm of nuclear security involves many parallel efforts in nonproliferation (verification of treaty compliance, monitoring for smuggling, proper storage and transportation of nuclear materials), cyber security, minimizing stocks of weaponizable materials, disaster response training, and nuclear forensics. All of these efforts have been continually improving, but there was a gap regarding the ability of the United States (US) to coordinate and respond to a nuclear incident, especially with the technical portion of nuclear forensics: characterization and analysis. After all, the first textbook on the topic was published in 2005 [9]. In 2006, the US Department of Homeland Security (DHS) founded the National Technical Nuclear Forensics Center (NTNFC) within the Domestic Nuclear Detection Office (DNDO). The mission of the NTNFC is to establish a robust nuclear forensics capability to attribute radioactive materials with demonstrable proof.

There are many fields that contribute to the nuclear forensics capability, such as radiochemical separations, material collection techniques, improving detector technology, material library development, and identifying forensic signatures. These needs vary based on whether the material being collected is post-detonation (e.g., bomb debris) or pre-detonation (e.g., SNF). In the pre-detonation realm, this project focuses on statistical methods to identify correlated material characteristics, which can lead to new forensic signatures.

1.1 Motivation

Nuclear forensics is an important aspect of deterring nuclear terrorism even though it is not, at first glance, thought to be preventative nuclear security. The most common defense of the field is that nuclear forensics capability deters state actors, not terrorist organizations. While it is true that a strong capability encourages governments to be more active in prevention of nuclear terrorism, it can also deter the terrorist organizations as well by increasing their chances of failure. Less destructive success tends to be more valued than high-risk mass destruction. In addition to influencing governments and making nuclear terrorism higher risk for organizations, nuclear forensics can assist in cutting off certain suppliers of nuclear materials or technologies (e.g., nuclear specialists that are only involved for financial reasons, access to state suppliers). Shutting off the sources builds a concrete barrier to nuclear terrorism. Therefore, nuclear forensics is considered impede this form of terrorism in both tangible and abstract ways [8].

Following the prevention value of nuclear forensics, it is important to understand the process of the technical portion of the investigation and how that can be improved. In the event of a nuclear incident, such as the retrieval of stolen special nuclear material (SNM) or the detonation of a dirty bomb, it is necessary to learn as much as possible about the source of the materials in a timely manner. In the case of non-detonated SNM, knowing the reactor parameters that produced it can point investigators in the right direction

in order to determine the chain of custody of the interdicted material. Determining these parameters (e.g., reactor type, cooling time, burnup) requires first characterizing and calculating certain isotopic ratios, chemical compounds, or trace elements. Both radiological methods (e.g., gamma spectroscopy) and ionization methods (e.g., mass spectrometry) measure these quantities. Although both measurement techniques have a multitude of techniques within them and thus varying strengths and weaknesses, the main tradeoff is between time/cost and amount of information gained.

The results of these analytic techniques are then compared against existing databases to obtain the desired reactor parameters. These databases are highly multidimensional, and furthermore, are rife with missing data entries and inconsistent uncertainties. Direct comparison between measurement results and a database therefore may not yield accurate results. Thus, computational techniques have been developed by nuclear engineers to calculate the parameters relevant to nuclear forensics analysis [1, 17, 18]. Another approach with the uniqueness of requiring minimal domain knowledge is the use of statistical methods via machine learning algorithms to predict these characteristics or values [2, 6, 7, 10, 11, 12, 15]. These algorithms can create a model using the database entries that enables "filling between the lines" of its entries. Additionally, having a machine-learned model may overcome the above challenges of multidimensionality, missing data, and irregular uncertainty.

1.1.1 Needs in Nuclear Forensics

The process of technical nuclear forensics includes the analysis and interpretation of nuclear material to determine its history, whether that be intercepted SNF, uranium ore concentrate (UOC), or the debris from an exploded nuclear device. After the technical portion is complete, intelligence data can be used to aid in material attribution; this is the overall goal of nuclear forensics.

After a nuclear incident, the material or debris is sampled and evaluated through many techniques that provide the following information: material structure, chemical and elemental compositions, and radioisotopic compositions and/or ratios. These measurements or ratios comprise the forensic signatures of the sample in question. These signatures can be analyzed with specific domain knowledge; for example, UOC will have trace elements depending on where it was mined from. They can also be analyzed against a forensics database in the case of SNF.

Measurement needs and techniques vary vastly depending on the material, as does the type of signature. This study focuses on non-detonated materials, specifically, SNF. It is important to determine if some intercepted SNF is from an undisclosed reactor or a commercial fuel cycle to attribute it to an entity or state. This is typically done by obtaining select chemical and elemental signatures and isotopic ratios, and comparing these measurements to those in an existing forensics database of reference SNF. The signatures for SNF correlate to characteristics that can, in a best case scenario, point

to the exact reactor from which the fuel was intercepted. The reactor parameters of interest are the reactor type, fuel type and enrichment at beginning of irradiation, cooling time, burnup [1, 17, 18].

The current and future work of this study are designed based on two primary needs to bolster the US nuclear forensics capability: forensics databases are imperfect, and our best measurement techniques are not always feasible in an emergency scenario. It is proposed that using a machine-learned model may be able to combat these issues; this is introduced next in Section 1.1.2.

SNF database imperfection is three-fold. First, forensics databases are kept by individual countries, and thus do not include a measurements from other reactor technologies around the world. Additionally, a given database will have widely varying uncertainty depending where and on what instrument the material was measured. Lastly, some fields in these databases have missing entries, which presents issues with matching and characterizing SNF based on interpolating between entries. cite

The second broad need within the forensics community is post-incident rapid characterization. A lofty goal would be to develop methods that provide instantaneous information, reliable enough to guide an investigation (e.g., within 24 hours). In the case of SNF, it takes weeks in a lab to measure isotopes via advanced (cooled detector) gamma spectroscopy and mass spectrometry equipment. A handheld detector that measures gamma spectra could provide the fast measurements to calculate isotopic ratios for the above-

mentioned fuel parameters of interest. However, while this nondestructive analysis is rapid, it is also difficult to evaluate because of the presence of overlapping peaks and the fact that uncertainties differ significantly because of the detector reponse, environment, storage, electronics, etc. Broadly speaking, gamma spectra give less information at a higher uncertainty than the near-perfect results of some destructive mass spectrometry techniques, like TIMS.

find MS technique and citation

1.1.2 Contribution of Statistical Methods

As previously mentioned, there are two main issues that are being addressed for forensics of SNF: database issues and speed of characterization. Many have begun considering computational approaches to nuclear forensics problems, such as the INDEPTH tool for inverse depletion and decay analysis [1, 17, 18]. This tool uses an iterative optimization method involving many forward simulations to obtain reactor parameters of interest given some initial values.

Another approach utilizes artificial intelligence to solve nuclear forensics problems, such as implementing searching algorithms for the database comparison step [3] and machine learning for determining reactor parameters from SNF characteristics [2, 6, 7, 10, 11, 12, 15]. A variety of statistical and machine learning tools have been used to characterize spent fuel by predicting categories or labels (e.g., reactor type, fuel type) as well as predicting values (e.g., burnup, initial enrichment, cooling time). The former uses classification

algorithms and the latter uses regression algorithms, many of which can be altered to perform both classification and regression.

This work evaluates to what degree statistical methods will be able to predict reactor parameters with respect to the type of training data used. There is some promising work discussed in Section 2.4 that shows certain applications of machine learning can provide an additional tool for solving the forensics problem, both qualitatively (for visualization) and quantitatively (for prediction).

1.2 Methodology

As previously mentioned, the typical workflow of the technical portion of a forensics investigation is to take measurements of an unknown material and compare those measurements to databases filled with previously measured standard materials. As this work focuses on SNF, these measurements are elemental, chemical, and radiological in nature. Because creating databases from real measurements to represent SNF from reactor technologies from around the world is impossible, the database in this study will be created from high-fidelity simulations via Oak Ridge Isotope GENeration (ORIGEN) [14] within the SCALE code system [13] for modeling and simulation.

In the simulation and statistical learning paradigm, we need to determine how much information to what quality is needed to train a machine-learned model; the model must give appropriate predictions of reactor parameters

given a set of measurements from a test sample of interdicted SNF. Thus, the space that the training set encompasses must be chosen carefully so as to represent the typical scenario for stolen SNF.

The next step is to choose an algorithm that performs statistical learning. Statistical learners have varied strengths and weaknesses based on what is being predicted and how they implement optimization. Chosen for this study are simple regression algorithms for burnup prediction: nearest neighbor and ridge regression. For comparison, support vector regression is used because it is known to handle highly dimensional data sets well. These algorithms are introduced in .

After the training is complete, the results of each models' predictions must be evaluated. Typically, a test set is used to compare against the model created from the training set. The testing error can therefore be tabulated with respect to various specifications such as the training set size, number of features, or algorithm parameters (regularization terms, etc). These results are broadly known as diagnostic plots and show if the algorithms' predictions are due to good performance or bad fitting.

After the models are evaluated using machine learning best practices, it will be important to compare them both against each other and against other computational forensics parameter methods. Thus, a Bayesian approach from the field of inverse problem theory will be used to give the probability density of the predictions so that the statistically generated predictions can be evaluated directly against other solutions, such as optimization-based

methods or direct computations.

Next, information reduction (within the training and/or testing data sets) must be investigated to extend this workflow to mimic that of the real world. The primary example investigated here is the reduction of information quality via gamma ray detectors, as they can provide fast results. If an algorithm could overcome the limitations of gamma detection and still provide useful results, this would warrant further studies and perhaps be field-applicable.

Thus, ultimately, the goal is to answer the question *How does the ability to determine forensic-relevant spent nuclear fuel attributes degrade as less information is available?*.

1.3 Goals

The main purpose of this work is to evaluate the utility of statistical methods as an approach to determine nuclear forensics-relevant quantities as less information is available. Machine learning algorithms will be used to train models to provide these values (e.g., reactor type, time since irradiation, burnup) from the available information. The training data will be simulated using the SCALE 6.2 code suite [13], which will provide an array of nuclide concentrations as the features (X) and the parameters of interest (y) are provided from the simulation inputs. Information reduction will be carried out using computationally generated gamma spectra; the radionuclide

concentrations from the simulations can be converted into gamma energies, which then undergo a detector response calculation to represent real gamma spectra as closely as possible. Machine learning best practices will be used to evaluate the performance of the chosen algorithms, and inverse problem theory will be used to provide an interval of confidence in the model predictions.

The necessary background is covered in Chapter 2. First, an introduction to the broader field of nuclear forensics is in Section 2.1 to place this work in the context of the technical mission areas. After that, a short discussion of the field of machine learning, the algorithms used, and validation methods are in Section 2.2. Section 2.3 includes information about the codes used to generate the training data, via fuel cycle simulation, detector response function, and isotope identification of gamma spectra. Lastly, a review of statistical methods being used in studies of forensics analysis is covered next in Section 2.4.

After the existing work is discussed and the gap that this work will fill is identified, the methodology is introduced next in Chapter 3. First, the experimental components and methods for validation are shown in Section 3.1, followed by demonstrations of the methodology in Section 3.2. Chapter 4 summarizes the official thesis research proposal and corresponding hypotheses. Finally, future directions and alternative directions are identified in Chapter 5.

2 BACKGROUND AND LITERATURE REVIEW

This chapter provides a background and literature review of the necessary components for this project. Section 2.1 outlines the broader field of technical nuclear forensics, with a focus on the area that motivates this project. Section 2.2 introduces the field of machine learning for an uninitiated audience, covers the relevant algorithms, and presents the methods field practitioners use for validation. Next, Section 2.3 covers the computational methods used to generate the training data for the machine learning input. Finally, the marriage of Sections 2.1, 2.2, and 2.3, is presented in Section 2.4, which is a review of previous work applying statistical methods to the nuclear forensics analysis of pre-detonated nuclear materials.

2.1 Nuclear Forensics

Nuclear forensics comprises a large part of an investigation into a nuclear incident, such as interdicted nuclear material or the detonation of a weapon containing radioactive components. The forensics portion of the investigation encompasses both the analysis of nuclear material and/or related paraphrenalia as well as the interpretation of these results to establish nuclear material provenance. The former has many technical aspects, relying on a range of nuclear science and chemistry. The latter involves intelligence and political considerations of the material analyses for attribution. This review

will only consider the technical portion of the nuclear forensics workflow.

Furthermore, the technical programs researching improvements to the US's nuclear forensics capabilities are split between the type of material being investigated. The analysis of irradiated debris from a weapon has different collection and measurement requirements than recovered SNF from a commercial reactor. This separates the field into post-detonation and pre-detonation nuclear forensics. While both are discussed in Sections 2.1.1.1 and 2.1.1.2, respectively, there is more focus on pre-detonation topics since this work is based on SNF.

Additionally, nuclear forensics is a traditional inverse problem, which has been well documented in mathematics and many scientific disciplines. Understanding inverse problem theory can help systematically define both the solution methods and their limitations. An introduction to the topic as well as its application to nuclear forensics is discussed in Section 2.1.2.

2.1.1 Types of Nuclear Forensics Investigations

2.1.1.1 Post-Detonation

Post-det

2.1.1.2 Pre-Detonation

Pre-det

2.1.2 Nuclear Forensics as an Inverse Problem

More inverse intro here.

As outlined in a textbook on the formal approach to inverse problem theory [16], the study of a typical physical system encompasses three areas:

1. Model parameterization
2. Forward problem predicting measurement values given model parameters
3. Inverse problem predicting model parameters given measurement values

First, this shows that it is important to consider the parameters that comprise a model; this is denoted as the *model space*. This is not every measurable quantity; domain knowledge is necessary to determine the model space. In the nuclear forensics context for spent nuclear fuel, this would consist of, e.g., several isotopic ratios because they are known to have a relationship with the reactor parameters that created the fuel of interest. Second, understanding the physical system also requires an understanding of the forward problem: predicting how a certain set of model parameters will affect the resulting measurements. The breadth of these end measurements provides the *data space*, which are all the conceivable results of a given forward problem. So for spent nuclear fuel this would be, perhaps, the range of isotopic ratios typical of a commercial reactor. Lastly, the inverse problem is statistical in nature: given some solution, there is a probability that the data measured is caused by some value(s) of a model parameter. Additionally, including measurement uncertainties broadens the linear model to a probability density of the parameters. The opposite is also true in

the forward case: including parameter uncertainties broadens the forward problem results to a probability density of the potential measurement values.

In this way, we can define some probability that the answer is correct, given a set of measurements and their uncertainty. Inverse problem theory states that this follows the general form of Bayes' theorem, which is commonly expressed as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.1}$$

where A and B are events, $P(A)$ and $P(B)$ are the probabilities that events A and B will occur, respectively, $P(B|A)$ is the prior probability that event B will occur given a known result for A, and $P(A|B)$ is the posterior probability that event A will occur given a known result for B.

This is can be mapped easily to the inverse physical system problem scenario. A would represent an occurence of a parameter in the model space, and B would represent the measurement of some value. Thus, $P(A)$ is the probability of a parameter existing without any knowledge of B. This is known as the likelihood, usually given by some theory about the system. $P(B)$ is the probability of some measurement existing without any knowledge of A. This is known as the marginal likelihood, which is some homogeneous concept for the potential measurements that could be made (this only serves to scale to absolute probabilities and does not affect the relative probabilities). The prior probability $P(B|A)$ is the chance that a measurement is observed from a given parameter, representing the forward

problem. Lastly, the posterior probability is the chance of some parameter existing given some measurement, representing the inverse problem solution [16]. It may be more intuitive to consider the conceptual version of Bayes' theorem below.

$$Posterior = \frac{Prior * Likelihood}{Marginal\ Likelihood} \tag{2.2}$$

A discussion of how these values are obtained takes place in Section 2.2.3.

This framework is helpful for an experiment that intends to compare different methods for calculating the posterior probability of a system given some measurements. In the nuclear forensics context of pre-detonated materials, this would be a a set of probabilities for different parameters of interest, e.g., reactor type, burnup, cooling time, and enrichment of some interdicted spent fuel. The prior probabilities are obtained by a large set of forward problems, e.g., a database of spent fuel recipes and parameters. The likelihoods are obtained in differing ways. One method is expert-elicited values. Another is a predicted model from some theory or previously known relationship, e.g., empirical relations between isotopic ratios and certain reactor parameters.

2.2 Machine Learning

Machine learning is a sub-field of artificial intelligence (AI) within the broad category of computer science. The goal of AI is to create computer systems that respond to their environment according to some set of criteria or goal. For example, self-driving vehicles have computers on board that learn to avoid curbs and humans. It is common knowledge that the use of AI has been expanding at a rapid rate in recent years. News stories of major tech companies' AI advancements are frequent and news articles abound with data on which jobs will be replaced with AI in the near future.

While its use has been increasing in the commercial sector, there is also much anecdotal evidence to support the existence of a rapid increase of AI use in academic research across many disciplines beyond robotics. AI systems have been used in detection (e.g., fraud or spam), medical diagnostics, user analysis (e.g., Netflix ratings), and a host of scientific disciplines that have increasing amounts of multivariate data.

Much of the recent advances to the field of AI have occured in the statistical realm, which forgoes domain knowledge in favor of large data sets. Thus, machine learning and statistical learning has become somewhat of a separate field. Machine learning research focuses on the underlying algorithms using mathematical optimization, methods for pattern recognition, and computational statistics. As an application, however, this study is not concerned with computational time, but rather the ability to correctly

could explain more, cite changing science of ML

predict values and categories relevant to the nuclear forensics mission. This restricts the relevancy of the algorithms to the underlying theory and its impact on the resulting model's accuracy.

Machine learning algorithms can be separated into two main categories: unsupervised and supervised learning. The former groups or interprets a set of input data, predicting patterns or structures. The latter includes both the input and output data, enabling the trained model to predict future outputs. Broadly speaking, the unsupervised learning algorithms are designed for clustering data sets or dimensionality reduction (i.e., determining some subset or linear combination of features most relevant to the input data) of data sets. Supervised learning algorithms predict both discrete and continuous values via classification and regression, respectively. Some algorithms can perform both classification and regression, and neural networks can even be modified to perform either supervised or unsupervised learning. Additionally, various algorithms can be strung together, which is referred to as *ensemble methods*. One common way of doing this is performing deimensionality reduction prior to supervised learning.

As shown in Figure 2.1, a typical (supervised) machine learning workflow begins with a training data set, which has a number of *instances*, or rows of observations of . Each instance has some *attributes*, also referred to as *features*, and a label, which can be a categorical label or discrete/continuous values. The training data are then inserted into a statistical learner; this calculates some objective, minimizes or maximizes that objective, and

check that this common workflow qualifies as ensemble

provides some model. This model is typically evaluated using a test set that has the same set of attributes and labels (but different instances). The comparison of what the model predicts and the actual label gives the *testing error*. Depending on the performance and application, the model may need improvement from more training and/or some changes in the algorithm parameters. Once the model is performing well enough, it is finalized; then a user can provide a single instance and a value can be predicted from that.

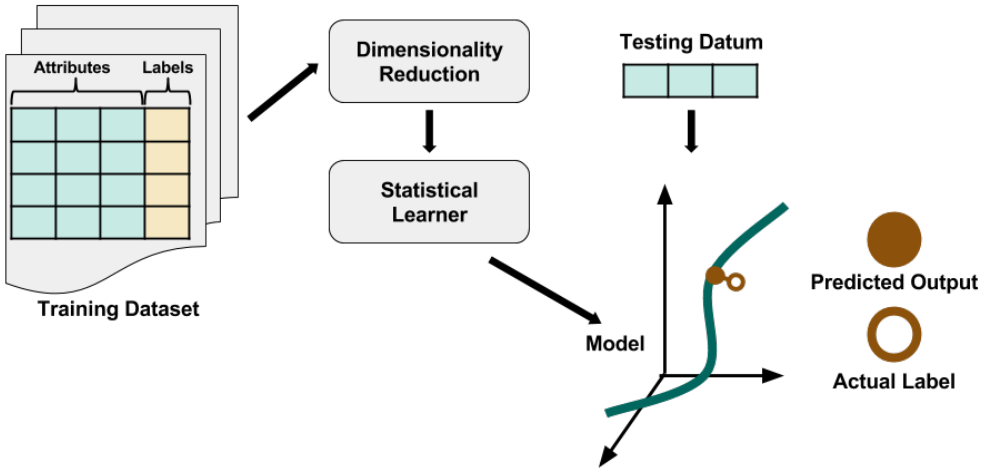

Figure 2.1: Schematic representing the workflow of a statistical learning regression algorithm.

This study performs both classification and regression tasks using supervised learning algorithms. Differences among the underlying mathematics of the algorithms do impact the performance of a machine-learned model. Therefore the algorithms used in this study will be discussed in Section 3.1.2.1. Additionally, evaluating algorithm performance is important and will

phrase

better

be covered in Section 2.2.2. Finally, robustly comparing different algorithms continues the previous discussion on inverse problem theory and is in Section 2.2.3.

2.2.1 Classification and Regression Algorithms

For relevant nuclear forensics predictions, both classification and regression algorithms must be used. For example, one may want to predict the reactor type based on some measurements (referred to as features) of spent fuel of an unknown source, and this would require a classification algorithm. Or perhaps the input fuel composition is relevant to an investigation on weapons intent, so a regression algorithm would be used to train a model based on some set of features. Since algorithm formulation impacts the resulting performance, they are discussed in detail below.

should add useful vocab: training data X and y, instances, features, etc

2.2.1.1 Linear Models

One of the simplest and most obvious methods of prediction is a linear model using a least-squares fit.

not sure about this organization

2.2.1.2 Nearest Neighbor Methods

Nearest neighrbor is

intro, alg math,

explain distance metrics

$$\hat{Y}(x) = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i \tag{2.3}$$

Nearest neighbor regression calculates a value based on the instance that is closest to it. The metrics for distance differ, but in this study, Euclidian distance was used. There is no learning in this regression, per se; the training set populates a space and the testing set is compared directly to that. [4]

2.2.1.3 Support Vector Machines

add schematics from pres...or diff pics with the math?

Support vector regression (SVR) is an extension of the popular classification algorithm, support vector machine (SVM). This algorithm was chosen because of its ability to handle highly dimensional data well, which in this study is approximately 300 features.

SVM classifies two classes by determining an optimal hyperplane, given by wx+b, between them. As seen in Figure ?, the algorithm evaluates the quality of the line that separates two classes by maximizing the width of the margin given the contraints surrounding the line. Some problems are not linearly separable, and thus a penalty term is introduced to allow for misclassifications. As shown in Figure ?, the algorithm then simultaneously minimizes the misclassifications while maximizing the margin.

This can be extended easily to multidimensional analysis via what is

called the *kernel trick*. First, using a nonlinear kernel function maps the data into higher dimensional feature space. Then the algorithm can find a linear separation in this space, as shown in Figure ?. Further, this can be upgraded from classification to SVR by doing similar math but instead minimizing the margin, as shown in Figure ?.

The kernel chosen for this study is the Gaussian radial basis function, shown below. This has two tuneable parameters, gamma and C. Gamma influences the width of influence of individual training instances, and strongly affects the fitting of the model. Low values correspond to underfitting because the instances have too large of a radius (low influence) and high values correspond to overfitting because the instances have a small radius (high influence).

The C parameter also affects the fitting of the model by allowing more or less support vectors, corresponding to more or less misclassification. A lower C smooths the surface of the model by allowing more misclassifications, whereas a higher C classifies more training examples by allowing fewer misclassifications. Too low and too high of a C can cause under- and overfitting, respoectively.

Since there is a tradeoff of fitting strength provided by both parameters, it is common to run the algorithm on a logarithmic grid from 10'-3 to 10'3 for each parameter. If plotted on a heatmap of accuracies given gamma and C, there will be a diagonal of ideal combinations that emerges. The lowest of these is usually chosen.

2.2.1.4 Artificial Neural Networks

dlllllll

2.2.2 Model Diagnostics

It is first important to determine if statistical methods can overcome the inherent database deficiencies. After that, the statistical methods must be considered in such a way as to represent a real-world scenario. Although mass spectrometry techniques provide extremely accurate isotopic information for analytical methods, they are time-consuming and more expensive. And although gamma spectroscopy can give extremely fast results cheaply, it only measures certain radiological signals and is influenced by many environmental factors, storage, and self-attenuation. As different machine learning algorithms and parameters are investigated, this work focuses on probing the amount of information required to obtain realistic results.

When plotting the training and testing error with respect to the number of instances, this is known as a learning curve. When plotting these errors with respect to the number or features or algorithm parameters, this is known as a validation curve.

bias-variance tradeoff

diagnostic and learning curves

insert example diagnostic plot?

2.2.3 Model Selection

All methods: Priors given by set of forward problems (input data space - ORIGEN sims) Likelihood given by model space, e.g.: Model params as determined from ML alg (related) MLE from bayesian inference Direct inversion of matrix A Expert-elicited params (i.e. initial guesses for INDEPTH? Or optimized ones?) Marginal likelihood only needed for absolute measures - doesnâĽ™t affect relative probabilities

2.3 Computational Methods

I'm a section on the computational methods!

2.3.1 Fuel Cycle Simulation

Short blurb on FC sim (capabilities enable other materials to be studied beyond SNF)

2.3.2 Data Modification

Perfect data -> real world data

2.3.2.1 Detector Response Functions

GADRAS?

Discuss methods, and potential options. Am I implementing for prelim?

2.3.2.2 Isotope Identification from Gamma Spectra

Same as above, but prob won't implement for prelim.

2.4 Applications of Statistical Methods to Nuclear Forensics Analysis

I'm the lit review section on previous applied ML work in the NF field.

2.4.1 Special Nuclear Materials Studied

The review on nf for the whole fuel cycle is useful here, perhaps. This is also important when I discuss my risk management section later.

2.4.2 Statistical Methods Employed

Very short details from lit review outline and success rates should be discussed here

3 METHODOLOGY AND DEMONSTRATION

This chapter first covers the methodology of the proposed work by introducing each experimental component in Section 3.1. Section 3.1.1 discusses how the training data is simulated for input to the machine learning algorithms. Section 3.1.2 is about how these will use the features and labels of the training data to statistically formulate a model for prediction of a new instance, which has only features and no label. The algorithms will be evaluated for accuracy and validated, as shown in Section 3.1.3.

Next is the demonstration of these experimental components, by first showing in Section 3.2.1 how the initial results guided the next steps for future experiments, discussed in Section 3.2.2. Steps taken towards model comparison (or even comparison against non-statistical methods) are shown in 3.2.3.

update me if necessary

3.1 Experimental Methodology

This work incorporates some methods and suggestions from previous work on the subject [2] regarding machine learning model performance with respect to information reduction, and expands upon it in two ways. The first is adding a different information reduction technique via applying a gamma spectroscopy detector response function (DRF) to the SNF nuclide recipes, which can calculate various spectra based on the types of gamma

detectors available to the forensics community. Secondly, a more advanced machine learning algorithm, support vector regression, is included so as to compare more complex models against the simpler models. A schematic of the workflow involving the experimental components is shown here in Figure 3.1. This section covers each of them as follows.

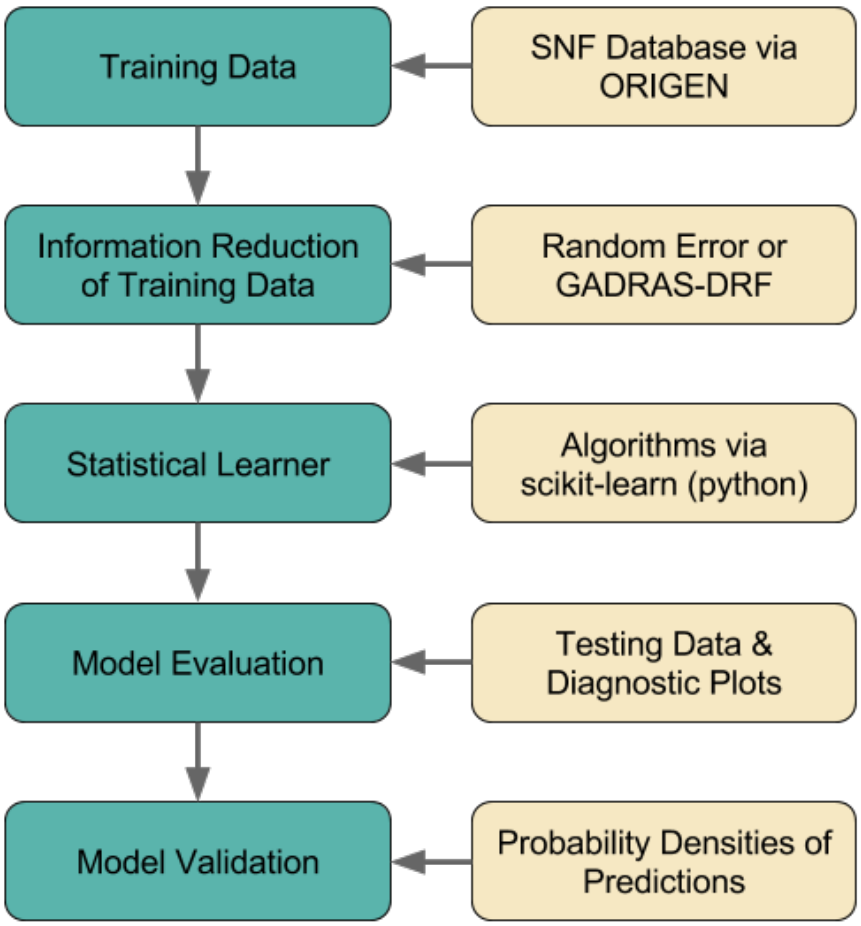

Figure 3.1: Methodology of the proposed experiment.

After the initial training data is simulated in Section 3.1.1.1, with a possible information reduction step in Section 3.1.1.2, it will be input to a statistical learner. While algorithm choice is discussed in Section 3.1.2.1, the main goal for these machine-learned models is to predict reactor parameters associated with some unknown SNF. This is shown in Section 3.1.2.2. To both understand the performance and validate the models, the results are then evaluated for over- or under-fitting, which is in Section 3.1.3.1. But validation is more than just making sure the models are properly fit to the data. Perhaps the training set was not representative of the actual data space, whereas other methods do not rely on the data space for results. So, lastly, comparison against other algorithms as well as other methods is described in Section 3.1.3.2.

3.1.1 Training Data

3.1.1.1 SNF Simulations

Because creating databases from real measurements to represent reactor technologies from around the world is impossible, the database in this study will be created from high-fidelity simulations via ORIGEN [14], an activation and depletion code within the SCALE 6.2 modeling and simulation suite [13]. Specifically, the ARP module of the activation and depletion code ORIGEN was used: ORIGEN-Automatic Rapid Processing (ORIGEN-ARP).

Of interest to an entity trying to create a weapon is partially irradiated

fuel if they have plutonium separations capabilities or any radioactive substance in the case of a dirty bomb. Addressing the former, a set of simulations of SNF at different burnups and cooling times will comprise the database. The training database is constructed by simulating the same training set space as described in Ref. [2].

The parameters of the training set are defined as follows. A smaller burnup than is typical for spent fuel from a commercial reactor is used in the previous work likely because stolen fuel pins for weapons use would not likely be at the end of their lifetime, as the plutonium of interest has decreased by then. A truly i.i.d. training set would go beyond this, but this is purely for demonstration with a single use case in mind.

show table of training set space

The previous work also used an external test set, designed to have values in between the trained values of burnup. This is implemented in this study but it is expected that cross-validation will better indicate the model performance. More specifically, using k-fold cross-validation is a common method to use in the application of machine learning to create more confidence in the resulting model.

show table of test set space

3.1.1.2 Information Reduction

In the simulation and statistical learning paradigm, we need to determine how much information to what quality is needed to train a machine-learned model; the model must give appropriate predictions of reactor parameters given a set of measurements from a test sample of interdicted SNF.

either explain or cite cross validation next

I am GADRAS-Detector Response Function (GADRAS-DRF) [5], yet another experimental component!

This study evaluates the impact of randomly introduced error of varying amounts on the ability of the algorithms to correctly predict the burnup. Thus, first investigated are the three algorithms with no error introduced, and next with the random errors applied uniformly to each nuclide vector. Since error in a nuclide vector is not random, in fact it is systematic and dependent on a number of known sources of uncertainty, the next study will introduce error by limiting the nuclides to only those that can be measured with a gamma spectrometer (future work).

3.1.2 Statistical Learning for Models

3.1.2.1 Algorithm Choice

Choice: Algorithm choice is usually based on what is being predicted and intuition regarding strengths and weaknesses of different optimizations. For a benchmarking excercise, some machine learning approaches here were chosen based on previous work [2]: nearest neighbor and ridge regression. This work will also extend to a more complex model via an algorithm that is known to handle highly dimensional data sets well: support vector regression. These algorithms are introduced in .

Many algorithms are developed on an assumption that the training set will be independent and identically distributed (i.i.d.). This is important so

that the model does not overvalue or overfit a certain area in the training space. The testing error can therefore be

3.1.2.2 Reactor Parameter Prediction

Discuss

3.1.3 Validation

Evaluation and Diagnostics: To obtain reliable models, one must both choose or create a training set carefully and study the impact of various algorithm parameters on the error. Many algorithms are developed on an assumption that the training set will be independent and identially distributed (i.i.d.). This is important so that the model does not overvalue or overfit a certain area in the training space. The testing error can therefore be tabulated with respect to training set size, number of features, or algorithm parameters (regularization terms, etc). The results are broadly known as diagnostic plots.

Alg Compare: inverse prob theory

As previously mentioned in the demonstration section, the validation was carried out as a comparison of accuracies of the predicted classes (reactor type,) and values (enrichment amount) to the classes and values in the predetermined test data set (discussed in).

However, because it is difficult to ensure consistently representative testing data, the accuracy of a learned model should not depend on only

add label

other classification tasks?

or burnup?

add label when

one testing set . The learned model's accuracy can be better evaluated as discussed in by using cross validation . Other additional evaluation methods will be discussed here as well.

In addition to evaluating a single learned model, it may be beneficial to compare models. As discussed in , there are three methods that will be used: comparison of receiver operating characteristics (ROC) curves, scatter plots, and pairwise t-tests.

Additionally, machine learning algorithms are heavily dependent on the inputs and parameters given to them, such as training set sizes, learning rates, regularization, etc. To evaluate the performance or tweak the model from an algorithm, diagnostic plots will be used. Learning and validation curves will indicate how the models are performing, initially both with respect to the testing error and the cross validation error. As previously mentioned, these two errors are to be compared to the training error to understand the prediction and generalization strength with respect to training set size and the algorithm parameters governing model complexity.

The learning curves were obtained as follows. For a given (randomly chosen) training set size between 15 and 100% of the total data set, several training and prediction rounds were performed. The repetition for obtaining the testing error is the same value as the k in k-fold cross validation. The testing error scenario averages the values of the obtained errors whereas the k-fold cross-validation performs this automatically. The validation curves were obtained as follows. For a given parameter, the value of the parameter

general knowledge or citation?

the algorithm validation section

should this be introduced in the background?

ROC curves, confidence intervals on error, confusion matrix if neces-

is varied and k training and prediction phases are completed, and their errors averaged. Again, for k-fold cross-validation, these errors are already averaged. The learning curves help determine if we are over- or under-training. The validation curves help determine the optimal way to be robust to over- and under-fitting.

1. Learning curve 2. Validation curve 3. Random error curve

Want to show the difference between test/train error plots and CV/train error plots. Determine some argument that perfers the latter. For the above categories in the validation section, can show plots directly next to each other (for 1, 2, 3) to hopefully show that cross validation provides better generalizability.

Options for comparison of algorithms: inverse bayesian stuff, comparing classification of 2 classes on same ROC plot with multiple ML systems, Scatter plots, Pairwise t-tests.

Discuss understanding confidence intervals in predictions.

3.1.3.1 Algorithm Evaluation

Discuss

3.1.3.2 Algorithm Comparison

Discuss

3.2 Experimental Demonstration

redo outline/-section titles here later

The first steps of this work include a loose replication of the previously mentioned work studying the effect of information reduction on predictions [2], and comparing this to the more complex algorithm chosen. This helped to establish some baseline expectations of reactor parameter prediction and how the different algorithms perform.

First discussed are the preliminary results in Section 3.2.1, which indicated the models were severely underfitted. Following that, the next step was to provide a larger, more diverse training set to the algorithms so they could predict better when faced with new instances. This is shown in Section 3.2.2.

Not sure about last section yet

3.2.1 Results

Some kind of results section here with validation, or just eval and conclusions.

3.2.2 Expanding the Training Set Space

This is what I did based on the results, and here are more results and validation. Dayman training/test set -> sfcompo-like sims and testing set

3.2.3 Comparing Different Models

Will I have prob density for prelim? Prob not?

4 RESEARCH PROPOSAL

This is what I'm gonna do.

4.1 Summary of Experiments

Summary

4.2 Hypotheses

I hypothesize!

5 FUTURE AND ALTERNATIVE WORK

This is what I could do, or might could do....

5.1 Future Work

This is how I will expand upon the experiments given the time. So many options.

5.2 Alternative Directions

This is the section on risk mitigation. Basically, say how UOC, Pu, or post-det could be covered with similar methodology.

REFERENCES

[1] Broadhead, Bryan L, and Charles F Weber. 2010. Validation of inverse methods applied to forensic analysis of spent fuel. In *Proceedings of the Institute of Nuclear Materials Management 51st Annual Meeting*. Baltimore, MD, USA. https://www.osti.gov/scitech/biblio/1001291.

[2] Dayman, Kenneth, and Steven Biegalski. 2013. Feasibility of fuel cycle characterization using multiple nuclide signatures. *Journal of Radioanalytical and Nuclear Chemistry* 296:195–201. http://link.springer.com/article/10.1007%2Fs10967-012-1987-4.

[3] Gey, Frederic, Chloe Reynolds, Ray Larson, and Electra Sutton. 2012. Nuclear forensics: A scientific search problem. In *Proceedings of the Lernen, Wissen, Adaption (Learning, Knowledge, Adaptation) Conference*. Dortmund, Germany. http://metadata.berkeley.edu/nuclear-forensics/Paper_9-12-12_lwa-2012-nuclear-forensics-scientific-search-problem_v7.pdf.

[4] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2001. *The Elements of Statistical Learning*. Springer Series in Statistics, New York, NY, USA: Springer New York Inc.

[5] Horne, Steven M., Gregory G Thoreson, Lisa A. Theisen, Dean J. Mitchell, Lee Harding, and Wendy A. Amai. 2014. Gamma De-

tector Response and Analysis Software - Detector Response Function (GADRAS-DRF). User's Manual, Sandia National Laboratories, Albuquerque, New Mexico, USA. Version 18.5; SAND2014-19465, http://www.osti.gov/scitech/servlets/purl/1166695.

[6] Jones, Andrew, Phillip Turner, Colin Zimmerman, and J.Y. Goulermas. 2014. Machine learning for classification and visualisation of radioactive substances for nuclear forensics. In *Techniques and Methods for Safeguards, Nonproliferation and Arms Control Verification Workshop*. Portland, Oregon. https://www.researchgate.net/publication/264352908_Machine_Learning_for_Classification_and_Visualisation_of_Radioactive_Substances_for_Nuclear_Forensics.

[7] Jones, Andrew E., Phillip Turner, Colin Zimmerman, and John Y. Goulermas. 2014. Classification of spent reactor fuel for nuclear forensics. *Analytical Chemistry* 86:5399–5405. http://pubs.acs.org/doi/ipdf/10.1021/ac5004757.

[8] May, Michael, Reza Abedin-Zadeh, Donald Barr, Albert Carnesale, Philip E. Coyle, Jay Davis, William Dorland, William Dunlop, Steve Fetter, Alexander Glaser, Ian D. Hutcheon, Francis Slakey, and Benn Tannenbaum. 2007. Nuclear Forensics: Role, State of the Art, and Program Needs. Tech. Rep., Joint Working Group of the American Physical Society and the American Association

for the Advancement of Science. https://www.aaas.org/report/nuclear-forensics-role-state-art-program-needs.

[9] Moody, K.J., P.M. Grant, and I.D. Hutcheon. 2005. *Nuclear Forensic Analysis*. 1st ed. Boca Raton, Florida, USA: CRC Press. https://books.google.com/books?id=Q9mgDnWoPLYC.

[10] Nicolaou, G. 2006. Determination of the origin of unknown irradiated nuclear fuel. *Journal of Environmental Radioactivity* 86:313–318. http://nuclear.ee.duth.gr/upload/A13%20%20%20identification.pdf.

[11] ———. 2009. Identification of unknown irradiated nuclear fuel through its fission product content. *Journal of Radioanalytical and Nuclear Chemistry* 279(2):503–508. http://link.springer.com./article/10.1007%2Fs10967-007-7300-x.

[12] ———. 2014. Discrimination of spent nuclear fuels in nuclear forensics through isotopic fingerprinting. *Annals of Nuclear Energy* 72:130–133. Technical Note, http://www.sciencedirect.com.ezproxy.library.wisc.edu/science/article/pii/S0306454914002308.

[13] Oak Ridge National Laboratory. 2016. SCALE: A Comprehensive Modeling and Simulation Suite for Nuclear Safety Analysis and Design. Code Suite, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA. Version 6.2.1, ORNL/TM-2005/39, Available from Radiation

Safety Information Computational Center as CCC-834, http://scale.ornl.gov.

[14] Rearden, B.T., and M.A. Jessee. 2016. Ch. 5 Depletion, Activation, and Spent Fuel Source Terms. In *SCALE Code System: User Documentation*, 5–1–5–263. Oak Ridge, Tennessee, USA: Oak Ridge National Laboratory. Version 6.2.1; ORNL/TM-2005/39, https://www.ornl.gov/sites/default/files/SCALE%20Code%20System.pdf.

[15] Robel, Martin, Michael J. Kristo, and Martin A. Heller. 2009. Nuclear forensic inferences using iterative multidimensional statistics. In *Proceedings of the Institute of Nuclear Materials Management 50th Annual Meeting*. Tuscon, AZ, USA: Institute of Nuclear Materials Management. LLNL-CONF-414001, https://e-reports-ext.llnl.gov/pdf/374432.pdf.

[16] Tarantola, Albert. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, chap. 1. The General Discrete Inverse Problem, 1–40. Philadelphia, Pennsylvania, USA: Society for Industrial and Applied Mathematics. http://epubs.siam.org/doi/pdf/10.1137/1.9780898717921.ch1.

[17] Weber, Charles F, Vladimir A Protopopescu, Michael H Ehinger, Alexander A Solodov, and Catherine E Romano. 2011. Inverse solutions in spectroscopic analysis with applications to problems in

global safeguards. In *Proceedings of the Institute of Nuclear Materials Management 52nd Annual Meeting*. Palm Desert, CA, USA. https://www.osti.gov/scitech/biblio/1031530.

[18] Weber, Chuck F, and Bryan L Broadhead. 2006. Inverse depletion/decay analysis using the scale code system. In *Transactions of the American Nuclear Society Winter Meeting*, vol. 95, 248–249. Albuquerque, NM, USA. Track 4: Nuclear and Criticality Safety Technologies.