SANDIA REPORT

SAND2024-15209
Printed Nov 2024

Radioisotope Identification with List-Mode Gamma Ray Data

A rigorous assessment on the value of temporal information applied to radioisotope identification

Lekha Patel, Efrain Gonzalez, Aaron Hill, Ryan Kamm

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185
Livermore, California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods

ABSTRACT

This work explores the potential of utilizing temporal data from gamma-ray detectors, known as list-mode data, to enhance radioisotope identification. Traditional identification methods, which rely on full gamma-ray spectrum analysis, often require long dwell times and struggle with "confuser" sources, or spectra with similarly spaced spectral peaks. We hypothesize that by leveraging the probabilistic nature of nuclear decay and the time-encoded information from decay sequences and interactions with surrounding materials, we can improve classification accuracy over static spectral analysis. This research rigorously examines the temporal content of list-mode data through exploratory data analysis via correlation discovery and information theory. We further propose a basic classification model that can utilize spectral or temporal data (or both) to determine if the incorporation of temporal information can improve radioisotope identification. The findings suggest that the temporal information present in list-mode gamma-ray data has merit and should be further investigated.

This page intentionally left blank.

ACKNOWLEDGMENT

We thank the LDRD office and SAIL LDRD Committee in supporting this work. This work was funded under LDRD Project #233708.

This page intentionally left blank.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

This page intentionally left blank.

ACRONYMS & DEFINITIONS

LDRD Lab Directed Research and Development

IPA Innovation for Proliferation Assessments

IA Investment Area

ARNIE Autonomous Reconfigurable Neural Intelligence At The Edge

N4PA Neuromorphic-computing for Proliferation Assessments

RIID Radioisotope Identification

LM-MLEM list-mode maximum likelihood expectation maximization

HMM Hidden Markov Models

MLE Maximum Likelihood Estimation

GADRAS Gamma Detector Response and Analysis Software

SNR Signal to Noise Ration

ROC Receiver Operating Characteristic

AUC Area Under Curve

MCA Multi-channel Analyzer

COTS Commercial Off The Shelf

CTMC continuous time Markov chain

This page intentionally left blank.

1. INTRODUCTION

Insight into proliferation activity is obtained using signatures of nuclear decay, typically gamma ray and/or neutron measurements, to identify radioisotopes. Radioisotope identification (RIID) is typically performed by analyzing gamma-ray spectra, sometimes in combination with neutron measurements. Current classification methods use peak-fitting or full gamma-ray spectrum analysis algorithms to identify isotopes, but these methods can require long dwell times to accurately classify radioisotopes in radiologically cluttered environments or weak signal scenarios [1, 2]. These methods also have difficulty with "confuser" sources, which have spectral peaks that are very close to each other, e.g., Pu-239 vs I-131 and U-238 vs Ra-226.

In the construction of a spectrum, time and sequence information is lost as gamma-ray counts are summed over time across energy bins—a fact that has been theoretically quantified [3]. Nuclear decay is a probabilistic process that follows well-characterized decay sequences and known mean decay times—information that is not present in a spectrum. Interactions of decay radiation with materials surrounding a source cause subsequent emissions with time-encoded information that can be attributed to geometry and materials. Utilizing the full temporal and sequential data from a detector, known as list-mode data, on its own or in combination with a spectrum, may improve methods for RIID, especially in the case of ambiguous spectra from confuser sources.

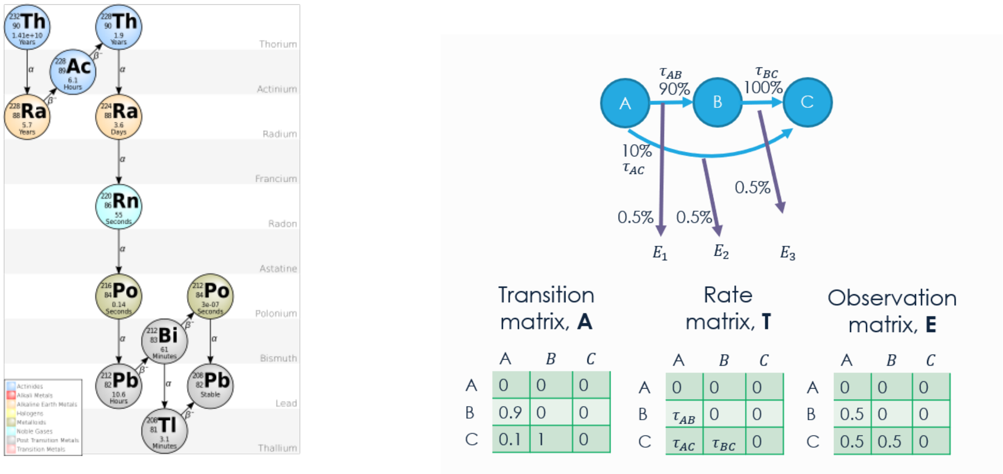
Some prior Sandia work has attempted to shed light on temporal information content in gamma-ray emissions, namely the ARNIE LDRD [4] and N4PA LDRD (current ongoing effort), both funded under the IPA IA, but it was not the focus of these efforts. Work from the BALDR project, funded under NA-22, applied machine learning models to the spectra-based RIID task, but admittedly still has trouble with confuser sources. Furthermore, list-mode data is used in Compton imaging with known isotopes, in algorithms such as LM-MLEM [5], and temporal based algorithms have been developed for fissile material characterization via active interrogation [6]. However, list-mode data has, to the author's knowledge, not been used successfully for passive gamma ray based RIID.

In this work, we take a rigorous approach to determine the nature of the temporal content of list-mode gamma-ray data. We begin by first addressing the Hidden Markov model hypothesis, in Section 3, that was central to the original LDRD idea submission (Idea # 24-0575). We then detail our datasets (Section 4) and data analysis efforts (Section 5) to define the nature of the temporal content of our datasets. This definition is performed through standard methods for correlation discovery and information theory analysis of the data. Lastly, we present a model for classification (Section 6) and accompanying results (Section 7) that can use spectral data alone, temporal data alone, or the fusion of the two, to determine if the temporal content of the data is beneficial to the task of RIID.

This page intentionally left blank.

2. MOTIVATION

Isotope branching ratios, decay modes and energies, and half-lives are readily accessible in several databases. For more complex radioisotope mixtures, there are many more metastable daughters and decay modes, potentially providing alternative signatures that can be captured in time-correlated list mode data. An example of one such decay chain is represented in Figure 2-1a. These decay chain representations look very similar to Markov chains or Markov processes. A Markov chain is a stochastic process that satisfies the Markov property. The Markov property states that the future state of the process depends only on the present state and not on the sequence of events that preceded it. Hidden Markov Models (HMMs) are statistical models used to represent systems that follow a Markov process with hidden states. They are particularly useful for modeling time series data where the system being modeled is assumed to be a Markov process with unobservable (hidden) states. It is possible to cast a decay chain into a Hidden Markov Model; see Figure 2-1b. In Section 3 we explore the application of this possibility to real radiation data and the physics of radiation sensing.



Transition matrix, **A**

	A	B	C
A	0	0	0
B	0.9	0	0
C	0.1	1	0

Rate matrix, **T**

	A	B	C
A	0	0	0
B	τ_{AB}	0	0
C	τ_{AC}	τ_{BC}	0

Observation matrix, **E**

	A	B	C
A	0	0	0
B	0.5	0	0
C	0.5	0.5	0

(a) Simple illustration of the decay chain associated with the radio isotope Thorium-232

(b) A toy example of a hidden Markov process model

Figure 2-1. On the surface, the representation of a radiological decay chain and a hidden Markov model are very similar.

Statistical theory postulates that in creating a gamma-ray spectrum using a stream of pulses coming from a detector and aggregating in time, information is being lost [3]. Therefore, using temporal gamma-ray data directly from the detector, before it is binned into a spectrum, can in principle be used to perform nuclear forensics and analysis with higher confidence and decrease identification/characterization thresholds. In Sections 5.3.1 and 5.3.2 we rigorously investigate this theory through Shannon Entropy analysis and Fisher Information quantification.

Further, nuclear decay is commonly modeled as a Poisson process, a stochastic process in which event counts are modeled via the Poisson distribution with a time dependency, exhibiting well-understood branching ratios and characteristic time scales for every decay. Gamma-ray emissions from a nuclear decay chain form fundamentally sequential information, and using only the integrated counts in a time interval disregards the sequence and time-of-arrival data. Using gamma-ray pulse measurements in a list-mode, real-time fashion, rather than a spectrum collected over time, allows temporal information in the data to contribute to the result, enabling isotopic identification to be made more rapidly than by using energy bin counts alone. In Section 6 we explore a classifier that can determine if the presence of temporal data from gamma-ray list-mode data can improve the accuracy of classifying radiological sources.

This work seeks to provide evidence and support to the temporal information content of list-mode gamma-ray data. We then ask, is the temporal information present useful to the task of radioisotope identification?

3. HIDDEN MARKOV MODELING OF LIST-MODE DATA

Radioactive decay is a stochastic process wherein unstable atomic nuclei spontaneously transform into more stable configurations, emitting radiation in the form of gamma rays, among many other forms. Focusing solely on gamma-ray emissions, radioactive decay can be stated to have the following properties:

1. Each decay event, i.e., particle emitted, occurs **independently** of others.
2. **Memoryless**: The probability of a decay event is constant over time, leading to an exponential distribution of interarrival times.
3. **Large Number of Atoms**: In practical quantities, radioactive sources contain an appreciable fraction of Avogadro's number ($\sim 6.022 \times 10^{23}$) of atoms that can be emitted during radioactive decay.

As such, radioisotope decay can be thought of as a continuous time Markov chain (CTMC) that transitions between different decay states and could be related to the measurement observations of time stamps and/or energies once a decay particle is observed. One may thus hypothesize that a Hidden Markov Model (HMM), which is able to relate observations to unobserved states, can be used to reliably identify these states and thus identify the radioisotope.

A Hidden Markov Model (HMM) [7, 8] is a statistical model where the system being studied is assumed to follow a Markov process with hidden, i.e., unobservable, states. These hidden states consist of a finite set $\{S_1, S_2, \ldots, S_N\}$ through which the system transitions over time. In continuous time [9], the likelihood of moving from one state to another is determined by the generator matrix A that defines the instantaneous transition rate from state S_i to state S_j. In addition to these transition rates, continuous time HMMs also include emission probabilities that define the likelihood of observing a particular output Y_t given the current hidden state X_t.

Let X_t denoting the hidden state at time t and Y_t the observed output, then the behavior of an HMM can be expressed mathematically as follows:

$$\mathbb{P}(X_t = S_j \mid X_{t-1} = S_i) = a_{ij}, \quad \mathbb{P}(Y_t = o_k \mid X_t = S_j) = b_j(o_k),$$

where a_{ij} denotes the elements of the generator matrix that yields the transition probability matrix over time t, $P_t = \exp(At)$, hosting the transition probabilities $p_{t,ij} = \mathbb{P}(X_t = S_j | X_0 = S_i$. Further, $b_j(o_k)$ are the emission probabilities that describe the likelihood of observing output o_k given that the system under observation is in state S_j.

We attempted to fit HMMs to energy data, ignoring the time stamp, with varying numbers of hidden states: 2, 3, 4 and 5. For example, for a 3-state HMM, the generator matrix A was specified as:

$$A = \begin{bmatrix} -a_{11} & a_{12} & a_{13} \\ a_{21} & -a_{22} & a_{23} \\ a_{31} & a_{32} & -a_{33} \end{bmatrix}$$

Emission distributions were defined using Poisson distributions with varying means, given that the measured energies are discrete. In all cases and data, however, each HMM failed to converge or provide sensible parameter estimates.

This result is not surprising when we take a step back and consider the nature of the physical phenomenon that we are attempting to model. Although HMMs are widely used in various fields for modeling systems that exhibit Markovian properties where the future state depends only on the current state, in the context of radioactive decay there are violations of the fundamental HMM assumptions that are required for an HMM to function (e.g., the stochastic nature of nuclear decay and practical observation limits) effectively rendering the HMMs impractical for analyzing list-mode gamma radiation data.

Implications of the HMM Assumption

An assumption of HMMs is violated when attempting to model a system composed of $N \approx 10^{23}$ independently decaying atoms using a single HMM chain. Each atom follows its own CTMC, independent of the others, as individual radioactive decay events are independent emissions from an unstable nucleus. When observing the overall decay process, we are capturing events from all atoms, rendering the use of a single HMM inadequate for modeling the collected data. To properly exploit the Markov property and relate each atom's isotope (hidden state) to the observed measurements, we would need to use on the order of 10^{23} independent HMMs, one for each atom. This usage would require a factorial HMM [10] to relate each atom's hidden state (isotope) to an observed event. However, inference and parameter estimation in this approach quickly becomes computationally intractable, as managing such a large number of independent chains would exceed the capacity of any modern computational resource.

Stochastic Nature of Nuclear Decay

Nuclear decay chains, which give rise to observable radiation signatures, are well understood. However, radioactive decay is a stochastic process, making it impossible to predict when a particular nucleus will decay, regardless of its age. For any given source-to-detector configuration, it is impossible to know whether a given emission related to a transition along a decay chain will be detected, and if it is, by what process. For example, a $\approx$ 1001 keV gamma ray from the decay of U-238 may only interact with a detector via Compton scatter, resulting in an observed energy deposition anywhere between 0 and $\approx$ 1001 keV.

Practical Observation Limitations

For any practical quantity of material of interest to the National Security Programs (NSP) Investment Area (IA), multiples of Avogadro's constant of decaying nuclei (e.g., hundreds of sextillions) will be under observation at once. The length of this observation is typically much shorter than even the short portions of a given decay chain. Using U-238 as an example, with a half-life of $\approx$ 4.5 billion years, it is clear that no practical NSP application should expect to make meaningful observations of decay chains of interest.

To illustrate the impracticality of using HMMs for this purpose, consider a thought experiment starting with the simplest case of a single atom in a 4π detector in a vacuum, where the detector can observe all emissions. As we remove ideal assumptions, the complexity of the system increases significantly:

1. Planar Detector: A finitely sized planar detector introduces solid angle loss, causing missed observations even with just one atom. Additionally, observations that do enter the plane of the detector will have a probability of being missed, as radiation interactions with the detector are probabilistic.
2. Terrestrial Environment: Introducing a terrestrial environment adds complexities such as Compton scattering, pair production, and emissions from natural sources, which modify and/or mask observations from the original source emission.
3. Multiple Atoms: Adding more than one atom (e.g., Avogadro numbers of them) results in emissions from atoms at different stages of the decay chain, further complicating the analysis.
4. Half-Life Considerations: A half-life is merely an observed length of time for which half the original atoms of a material have decayed, not a prescription of when emissions actually occur. Thus, emissions along the decay chain are completely unknowable until a perfect observation is made, which is not possible in any practical scenario.

Given these considerations, it is clear that Hidden Markov Models are not practical for the analysis of radiation list-mode data due to the quantity of independent decay events, computational intractability, stochastic nature of nuclear decay, and practical observation limitations. Consequently, whether HMMs can be leveraged to perform any useful classification of radioisotopes is not considered further here.

This page intentionally left blank.

4. DATASETS

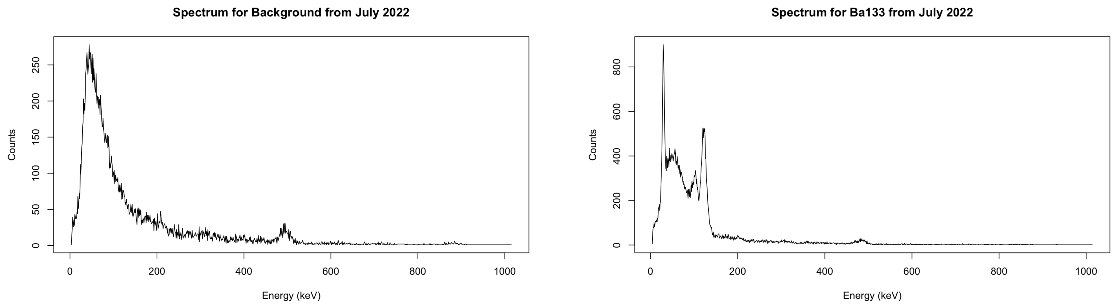
Our work in this report utilizes three different datasets collected in July 2022, April 2024, and August 2024. All these datasets come from radioactive sources collected with a 2” × 4” × 16” low resolution inorganic scintillator (Sodium Iodide (NaI) or Cesium Iodide (CsI)) connected to a COTS multichannel analyzer (MCA)—Canberra/Micron Osprey or digiBase. For all datasets, we did not utilize the first 100 seconds of data due to this period of time being reserved for experiment setup. All collections were performed with a stationary detector and a stationary source. In this section we detail each dataset’s sources with spectrum and any relevant details regarding the nature of the collection.

4.1. July 2022 Data

This dataset consists of five collections across four different sources, with the fifth collection being a combination of two of the sources. The collection was performed with a stationary detector and a stationary source. Collection periods were for approximately three hours. Sources used in this collection are as follows:

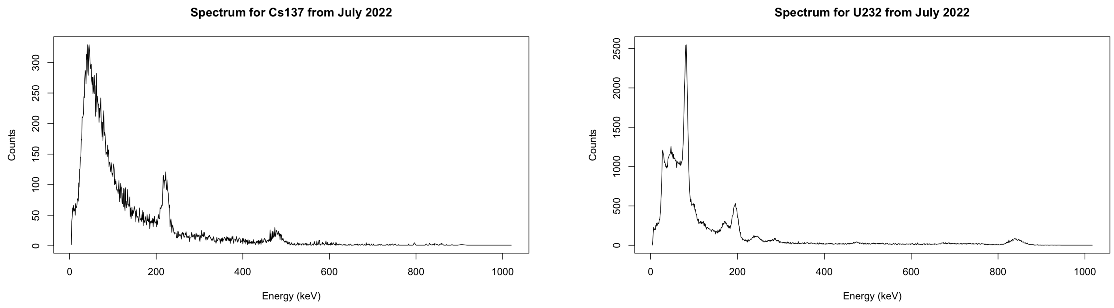
1. Cs-137
2. Y-88
3. Ba-133
4. U-232
5. Ba-133 + U-232

No calibration for drift was performed on this detector, and as such any data beyond 20 minutes was not used. Spectra based on 20s worth of events for each source are presented below in Figure 4-1.



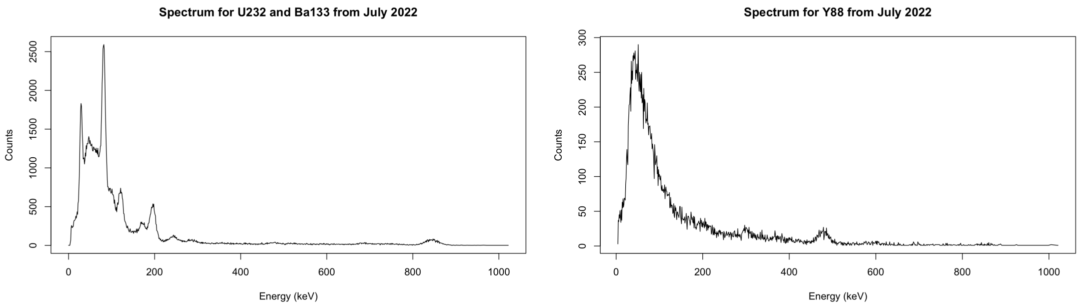
(a) Spectrum of Background from July 2022

(b) Spectrum of Ba133 from July 2022



(c) Spectrum of Cs137 from July 2022

(d) Spectrum of U232 from July 2022



(e) Spectrum of U232 & Ba133 from July 2022

(f) Spectrum of Y88 from July 2022

Figure 4-1. Figures 4-1a, 4-1b, 4-1c, 4-1d, 4-1e, and 4-1f display the spectrum associated with background, Ba133, Cs137, U232, the combination of U232 and Ba133, and Y88, respectively. The list-mode data associated with these plots was collected in July 2022.

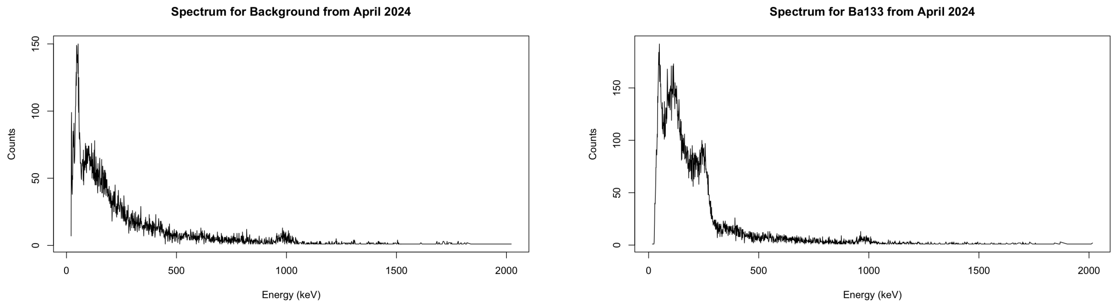
4.2. April 2024 Data

This dataset consists of three collections across two different sources, with the third collection being a combination of the two unique sources. The collection was performed with a stationary detector and a stationary source. Collection periods were not controlled. Sources used in this collection are as follows:

1. Ba-133
2. U-232

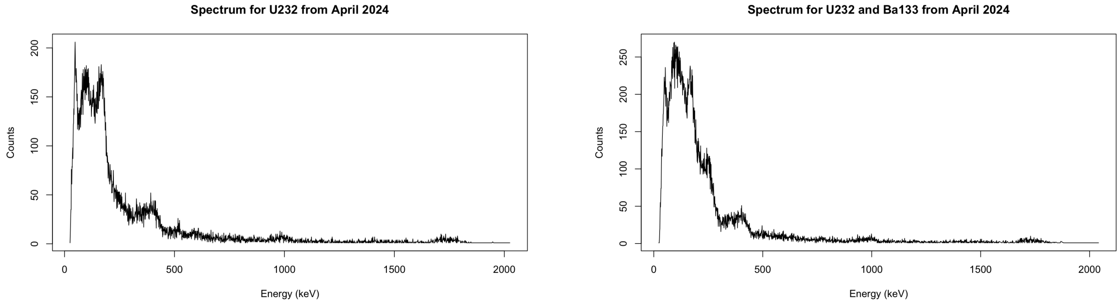
3. Ba-133 + U-232

Calibration of the detector was performed to account for drift from long dwell times. Spectra based on 20s worth of events for each source are presented below in Figure 4-2.



(a) Spectrum of Background from April 2024

(b) Spectrum of Ba133 from April 2024



(c) Spectrum of U232 from April 2024

(d) Spectrum of U232 & Ba133 from April 2024

Figure 4-2. Figures 4-2a, 4-2b, 4-2c, and 4-2d display the spectrum associated with background, Ba133, U232, and the combination of U232 and Ba133, respectively. The list-mode data associated with these plots was collected in April 2024.

4.3. August 2024 Data

This dataset was produced as a result of our initial results on the first two data sets. Because each source has a distinctly different gross count rate, we were concerned that the classifier may be exploiting the gross count difference as a discriminator. To test this hypothesis, this dataset placed the sources at different distances in an attempt to make the gross count rate of each collection the same. Sources used were as follows:

1. Cs-137
2. Y-88
3. Ba-133
4. Cs-137 + Ba-133

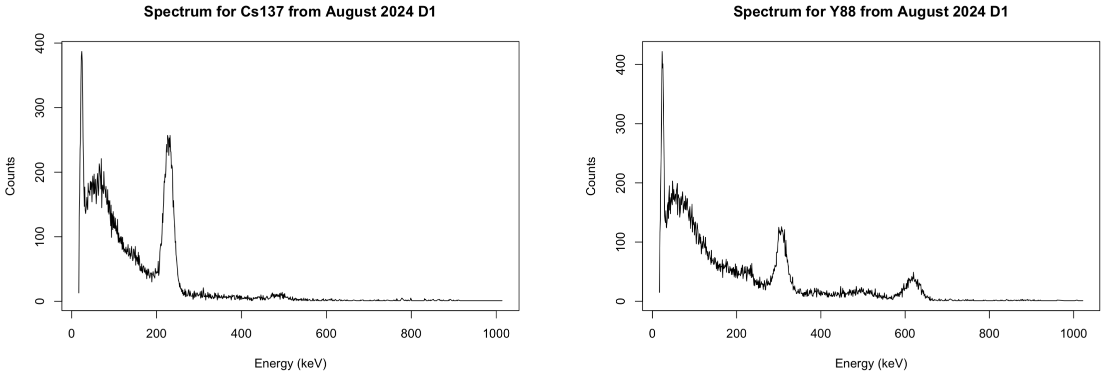
Table 4-1 details the distances each source was placed from the detector and the corresponding gross count rate achieved. Distance 1 refers to the set of distances to achieve a nominal count rate of between 1600 to 1800. That data is referred to as D1. Distance 2 refers to the set of distances to achieve a nominal count rate of between 1200 and 1400. That data is referred to as D2. For completeness, Table 4-2 makes record of the MCA settings used for this collection. Collection periods were not controlled. Spectra based on 20s worth of events for each source in datasets D1 and D2 can be found in Figures 4-3 and 4-4.

	Distance 1 (D1)		**Distance 2 (D2)**	
Source	Distance	Gross Count Rate	Distance	Gross Count Rate
Cs-137	10ft 2in	1598	15ft 2.5in	1181
Y-88	7ft 1.5in	1738	11ft 4in	1230
Ba-133	21ft 6in	1827	29ft	1381
Cs-137 & Ba-133	22ft 9in	1849	31ft	1350
Background		751		

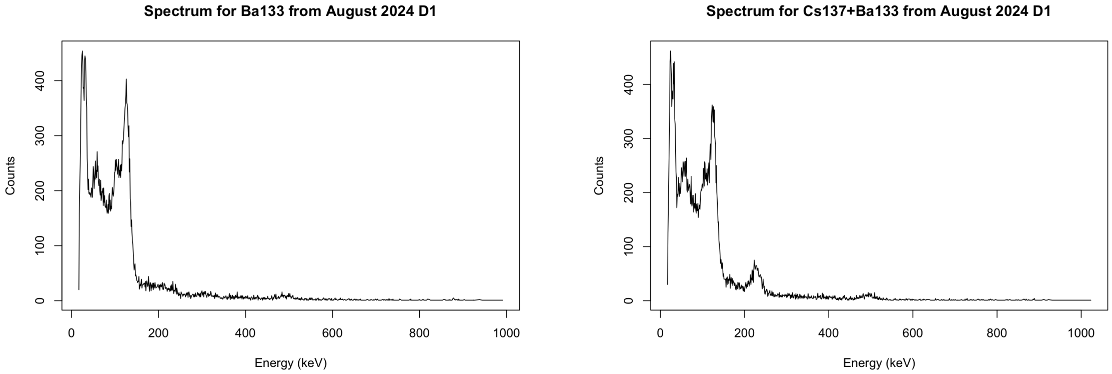
Table 4-1. Distances used for each source to achieve specific count rate targets.

Amplifier	Gain	0.4
	Shaping Time	2 μsec
ADC	Lower Level Disc	17
	Upper Level Disc	1023
High Voltage	Target	666

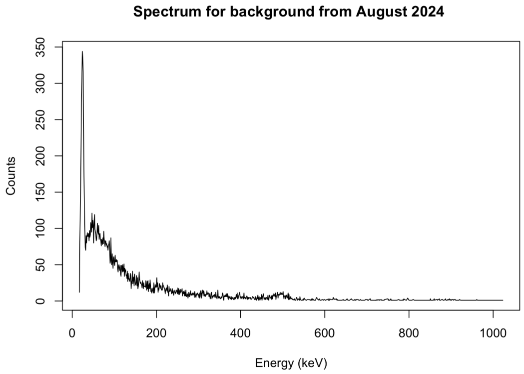
Table 4-2. MCA Settings for August 2024 dataset collection.



(a) Spectrum of Cs137 from August 2024 D1 **(b) Spectrum of Y88 from August 2024 D1**

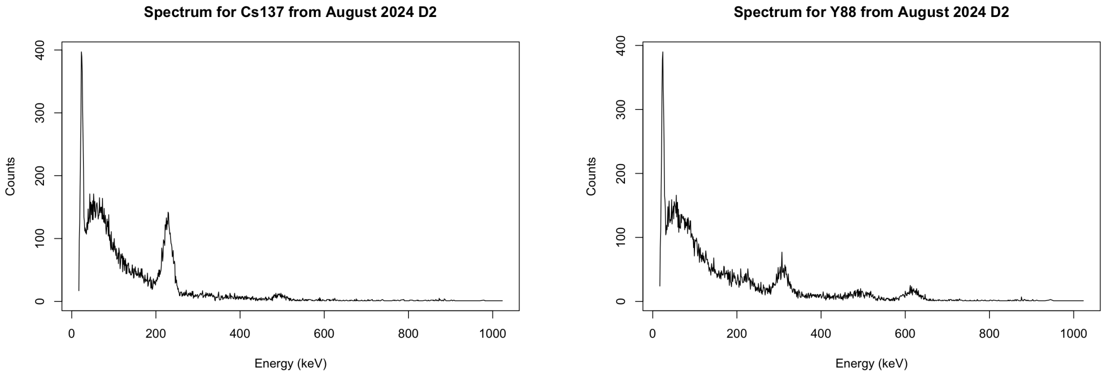


(c) Spectrum of Ba133 from August 2024 D1 **(d) Spectrum of Cs137 & Ba133 from August 2024 D1**

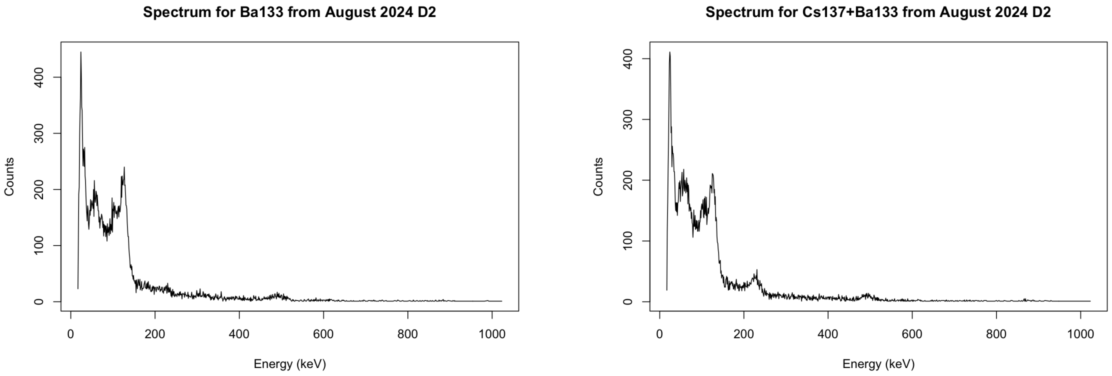


(e) Spectrum of Background from August 2024

Figure 4-3. Figures 4-3e, 4-3a, 4-3c, and 4-3d display the spectrum associated with Cs137, Y88, Ba133, the combination of Cs137 and Ba133, and background, respectively. The list-mode data associated with these plots was collected in August 2024 at D1.



(a) Spectrum of Cs137 from August 2024 D2 **(b) Spectrum of Y88 from August 2024 D2**



(c) Spectrum of Ba133 from August 2024 D2 **(d) Spectrum of Cs137 & Ba133 from August 2024 D2**

Figure 4-4. Figures 4-4b, 4-4a, 4-4c, and 4-4d display the spectrum associated with background, Y88, Cs137, Ba133, and the combination of Cs137 and Ba133, respectively. The list-mode data associated with these plots was collected in August 2024 at D2.

5. EXPLORATORY DATA ANALYSIS

In this section we explore the relationship between time and energy in the list-mode data that was collected. Section 5.1 shows the results of correlation analysis on the July 2022 and April 2024 datasets. Section 5.2 discusses the results of the distribution analysis that was conducted on the July 2022 dataset. Lastly, Section 5.3 discusses the results of the Shannon entropy and Fisher information analysis that was conducted on all datasets mentioned in Chapter 4. Except when stated otherwise, we used a 20s sample from each of the datasets in order to generate the results in this section.

5.1. Correlation

One way to analyze the relationship between two variables is by using the Pearson's correlation coefficient, which we will henceforth refer to as the correlation coefficient. The correlation coefficient is a unit-less measure of the strength and direction of the linear relationship between two variables [11]. It ranges in value from -1 to 1, with values that are larger in magnitude indicating that there is a strong linear relationship between the two variables. Due to the nature of the list-mode data that was collected, a strong linear relationship between interarrival time and energy was not expected. Table 5-1 shows that the Pearson's correlation coefficient between interarrival time and energy is nearly zero for every source, and therefore a linear relationship between energy and interarrival time does not exist in our datasets. An alternative measure of correlation is referred to as Spearman's rho. Spearman's rho is calculated in a similar way to Pearson's correlation coefficient except that it uses ranks and average ranks to arrive at a measure of the strength and direction of the monotonic relationship between two variables [12]. Again, due to the nature of our data, we did not expect there to be a strong monotonic relationship between interarrival time and energy. This hypothesis was validated by the results of this analysis that can be found in Table 5-1.

Additionally, since all of these datasets represent time-series data, we explored two correlation metrics that are specifically used for time-series data. The first metric is referred to as autocorrelation and is used to measure the strength and direction of the linear relationship between a variable at one moment in time and the same variable at another moment in time [13]. This metric can also be described as a measure of how well one can use a linear function to predict the $(j + k)^{th}$ element in the series when given the j^{th} element in the series. Just as before, the values for this metric fall between -1 and 1, where large values in magnitude imply a strong linear relationship. When using this metric we looked at interarrival times and energy separately. Since a linear relationship is not expected to exist between elements in the list-mode data, the autocorrelation value was hypothesized to be close to zero for both the interarrival times and the energy. The autocorrelation value between the j^{th} interarrival time and the $(j + k)^{th}$ interarrival

Relationship Between Interarrival Time & Energy				
July 2022	Pearson's Correlation	Spearman's Rho	Autocorrelation (Time, Energy)	Cross-Correlation
Background	-0.0028	-0.0007	(0.0151, 0.0083)	0.0147
Y88	-0.0019	-0.0004	(0.0142, 0.0136)	0.0156
Cs137	-0.0029	0.0009	(0.0115, 0.009)	0.0127
U232	-0.0037	-0.0017	(0.0071, 0.0068)	0.0059
Ba133	0.01	0.0085	(0.0077, 0.0089)	0.0118
U232 & Ba133	-0.0003	-0.0044	(0.004, 0.004)	0.0063
April 2024				
Background	-0.0092	-0.0218	(0.0196, 0.0154)	0.0166
U232	-0.0041	-0.0138	(0.0144, 0.012)	0.0145
Ba133	-0.0077	-0.0107	(0.013, 0.0105)	0.0123
U232 & Ba133	-0.0077	-0.0187	(0.0088, 0.0132)	0.0129

Table 5-1. The table displays the different correlation metrics that were used to analyze the list-mode datasets.

time was found to indicate a lack of correlation for every dataset. The maximum autocorrelation value is shown in Table 5-1, where the values of k that were considered were between 1 and 42. Similar results were obtained for the energy variable. Furthermore, cross-correlation was used to test whether there was a linear relationship between the series of interarrival times and the series of energies. The maximum cross-correlation between $x\left[t+k\right]$ and $y\left[t\right]$ is shown in Table 5-1, where $x\left[t+k\right]$ represents the interarrival time for the t^{th} event with a lag of $k \in \{-42, \ldots, 42\}$ and $y\left[t\right]$ represents the energy channel associated with the t^{th} event. All cross-correlation values indicate a lack of a linear relationship between the two variables.

The analysis presented above gives credence to the assumption that the energy channels and interarrival times are linearly independent of each other. It further confirms and extends the work from [4] that looked at similar correlation metrics under binned versions of radiation data. That same work also investigated Liquid State Machines (LSM, a class of algorithms well suited to exploit temporal correlations in data), by applying a genetic algorithm (GA) for optimization of the LSMs. The GA optimized out the very structure of the LSM that exploits temporal correlations, indicating the list-mode radiation data contains no temporal correlation.

5.2. Distributions

Although we were unable to establish the existence of a relationship between interarrival time and energy channel in the previous section, we investigated the energy and interarrival time distributions separately in order to evaluate whether or not there are differences in the distributions of each of these variables for each of the classes. This investigation was conducted by generating box plots from the 20s samples that we had used in the previous section. Box plots are a visual representation of the "five-number summary," which includes the first (Q_1), second (Q_2), and third quartile (Q_3) as well as the upper and lower fence. The second quartile is usually referred to as the median, and the upper and lower fences are calculated by using the following equations:

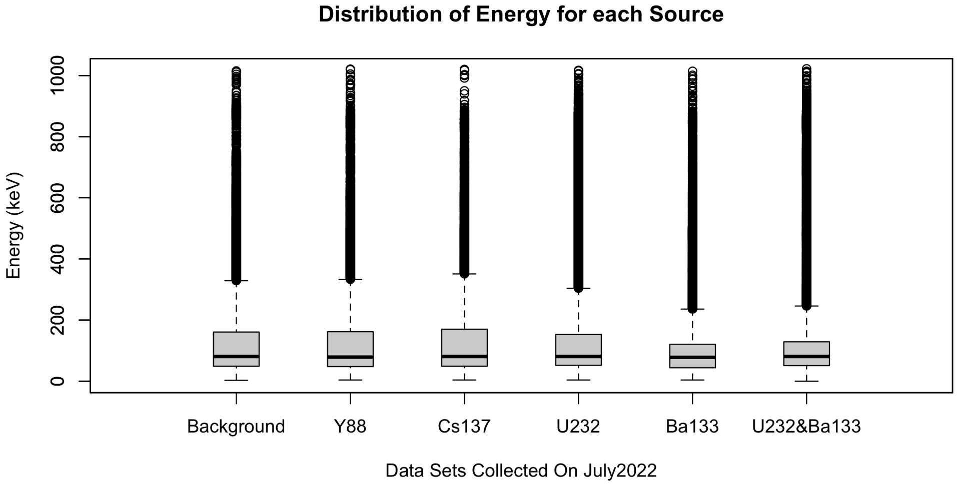
$$\text{upper fence: } Q_3 + 1.5 \times IQR \tag{5.1}$$

$$\text{lower fence: } Q_1 - 1.5 \times IQR \tag{5.2}$$

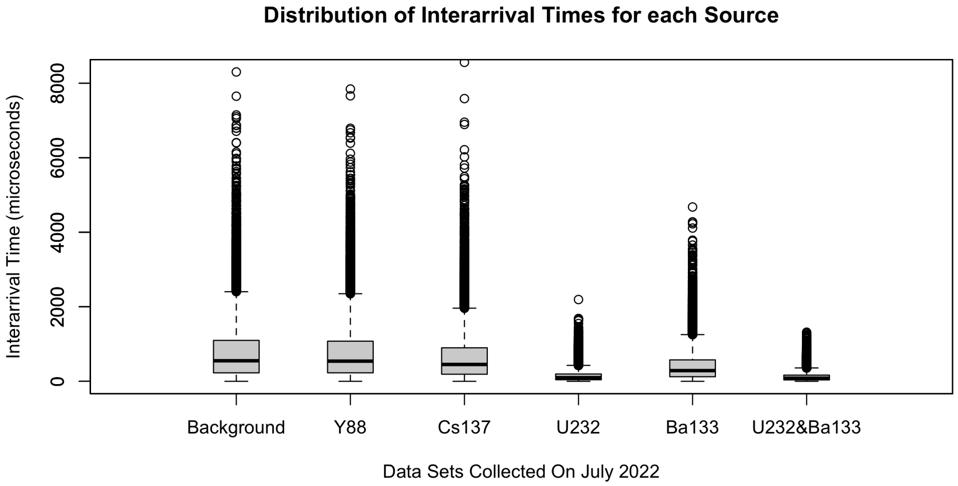
$$IQR = Q_3 - Q_1 \tag{5.3}$$

where IQR is called the interquartile range. Values outside of the fences are considered to be outliers. If the minimum value in the dataset is larger than the calculated value for the lower fence, then the lower fence value is set to the minimum value in the dataset. Likewise, if the maximum value in the dataset is smaller than the calculated value for the upper fence, then the value for the upper fence is set to the maximum value in the dataset. In the datasets that we dealt with, the former was a common occurrence because the smallest possible interarrival time is 1 microsecond. The results of this initial look at the distribution of the energy and interarrival times for each source are presented in Figure 5-1. As expected, the distribution of the energy channels varies based on the source dataset. This variation is most likely part of the reason that the spectrum is useful as a means for radioisotope identification. Additionally, a similar variability is present in the case of interarrival times, which establishes the potential utility of this additional variable when used in a model intended for radioisotope identification.

Since our work is primarily focused on leveraging the time dimension of list-mode data for radioisotope identification, we further studied the stability of the distribution of interarrival times by looking at how this distribution changes across 80 different 20s windows and how the distribution changes as the size of the sampling window changes. In the former case, each different 20s window was given an interval index, as can be seen in Figures 5-2a and 5-2b. In the latter case, 100 differently sized sampling windows were created by using the formula $20/d$, where d is an integer $d \in \{1, \ldots, 100\}$. Therefore the sampling window size ranged from 20s to 0.2s. Each of the differently sized sampling windows was given an interval length index based on the divisor used in the above formula. For example, an interval length index of 1 implies that the sampling window size was 20s, whereas an interval length index of 100 implies that the sampling window size was 0.2s. This phrasing was used in Figures 5-2c and 5-2d. As expected, the distribution of interarrival times did not change significantly across the 80 different 20s windows, but it did change as the size of the sampling window changed. By establishing the stability of the distribution of each dataset across several 20s windows we are able to provide evidence that shows that variability in the interarrival times across sources should be expected regardless of the 20s window that is chosen. These results can be seen in Figures A-1 and A-2 in Appendix A.1. Unlike in Figure 5-1, for this analysis, energy values less than or equal to 200 keV were filtered out as a way of removing some of the events that would typically be associated with background. It can be difficult to discern any possible changes in the distributions by looking at Figures A-1 and A-2, therefore summary statistics were calculated for each box plot. Figure 5-2 shows a few example plots of the median and interquartile range under these different cases, but the full list of generated plots is available in Appendix A.1. The y-axes of Figures 5-2a and 5-2b point to a lack of variability in the median and interquartile range as the sampling window is changed. However, the y-axes for Figures 5-2c and 5-2d indicate that as the length of the interval is varied the distribution of interarrival times varies more significantly than when the the sampling windows were changed.



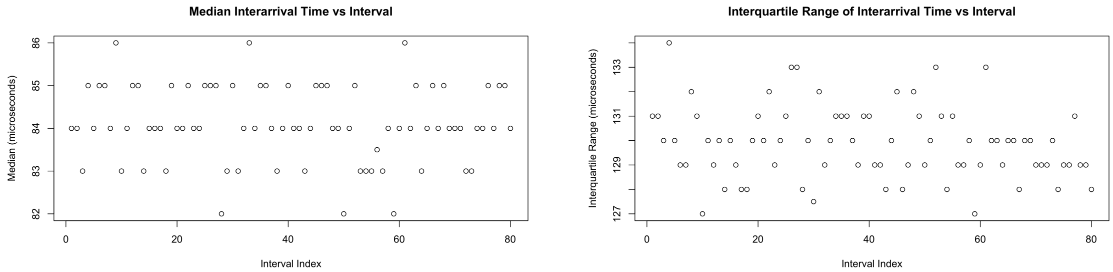
(a) Distribution of energy values for each source collected on July 2022



(b) Distribution of interarrival times for each source collected on July 2022

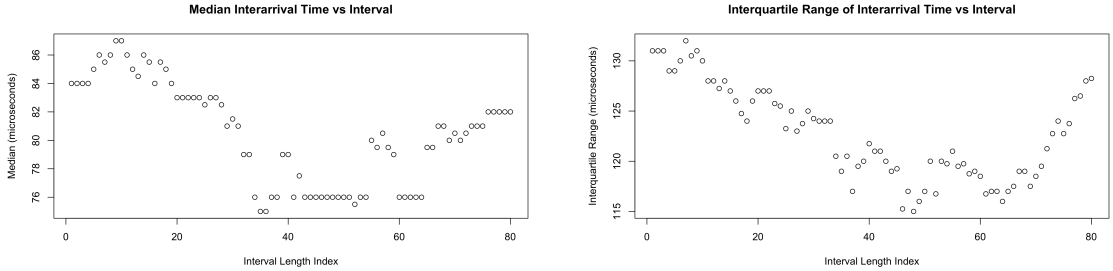
Figure 5-1. Figures 5-1a and 5-1b display the distribution of energy and interarrival time on a per source basis over the $20s$ window that was used for the correlation analysis presented in the previous subsection.

The observations discussed here for Figure 5-2 hold true for the other similar box plots found in Appendix A.1.



(a) Median of interarrival Times of U232 & Ba133 on July 2022

(b) IQR of interarrival Times of U232 & Ba133 on July 2022



(c) Median of interarrival Times of U232 & Ba133 on July 2022

(d) IQR of interarrival Times of U232 & Ba133 on July 2022

Figure 5-2. Figures 5-2a and 5-2c display the change in the median associated with the distribution of interarrival times for the combination of U232 and Ba133 as the sampling window changes and as the size of the sampling window changes, respectively. Figures 5-2b and 5-2d display the change in the interquartile range associated with the distribution of interarrival times for the combination of U232 and Ba133 as the sampling window changes and as the size of the sampling window changes, respectively. The list-mode data associated with these plots was collected in July 2022.

5.3. Understanding Information Present In List-Mode Data

Current state-of-the-art classifiers use spectra for radioisotope identification [14]. These spectra can be viewed as list-mode data that has been binned in time. In particular, the list-mode data over some interval of time is summarized into a histogram of counts per energy channel in order to generate the spectra. In order to gain insight into the effect of binning in time on the information present in list-mode data, we chose to study the impact of binning in time on two information metrics: Shannon entropy and Fisher information.

5.3.1. Background: Shannon Entropy

Stemming from information theory, Shannon entropy is used as a metric for quantifying the average level of uncertainty or information associated with the outcomes of a random variable.

Suppose X is a random variable with a probability mass function $p_X(x) = P(X = x)$ and let S be the set of possible values for X, then the Shannon entropy of X is defined as

$$H(X) = \mathbb{E}\left[-\log_2\left(p_X(x)\right)\right] = -\sum_{x\in S} p_X(x)\log_2\left(p_X(x)\right), \tag{5.4}$$

where here $\log_2$ refers to the logarithm with a base of 2 [15]. In order to understand this metric better, let us consider a couple of examples. First consider the extreme case where the $P(X = c) = 1$ and $c \in S \subseteq \mathbb{R}$, then the Shannon entropy of X is

$$H(X) = -\sum_{x\in S} p_X(x)\log_2\left(p_X(x)\right) = -p_X(c)\log_2\left(p_X(c)\right) = 0. \tag{5.5}$$

Therefore, the more certain one is of the value of our random variable, the less informative knowledge of that random variable becomes. As our second example, suppose that X is a random variable whose set of possible values are integers from $S = \{1, \ldots, n\}$; then it can be shown that the maximum Shannon entropy is obtained when $P(X = x) = \frac{1}{n}$ for $x \in S$. This result again implies that higher uncertainty about the value of the random variable results in a larger Shannon entropy value.

For our purposes, we wanted to determine the effect of binning on the information present in the data. Since we used list-mode data, which contains information on both time T and energy E, we needed to use the multivariate equation for Shannon entropy, which takes into account the joint distribution of both random variables. The multivariate equation for Shannon entropy is as follows

$$H(T, E) = H(T) + H(E|T), \tag{5.6}$$

where

$$H(E|T) = -\sum_{e,t\in E\times T} p_{E,T}(e,t)\log_2\left(\frac{p_{E,T}(e,t)}{p_T(t)}\right) [15]. \tag{5.7}$$

5.3.2. Background: Fisher Information

As can be determined by its definition found in equation (5.8), Fisher information is related to log-likelihood. The log-likelihood function for the binned and unbinned scenarios is defined in Section 6.2.4. Let $\ell(\boldsymbol{\theta})$ represent the likelihood function, then the Fisher information matrix looks as follows

$$I(\boldsymbol{\theta}) = -\mathbb{E}\left[\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}\right]. \tag{5.8}$$

Using equation (5.8) the Fisher information associated with the rate parameter of each source λ_k can be calculated in both the unbinned and binned log-likelihood function scenarios.

5.3.3. *Binning of List-Mode Data*

As stated previously, our goal was to determine the effects of binning on Shannon entropy and Fisher information. This goal required us to define a process for binning list-mode data progressively into larger time bins until we obtain the spectrum for the source that created the list-mode data. We consider the spectrum to be the fully-binned case and the original list-mode data to be the case where nothing has been binned.

Process for defining bins

The following shows how intermediary binning cases were defined:

1. Choose an integer number of bins b where $b \in \{1, \ldots, n\}$, 1 refers to the fully-binned case, and n refers to the original list-mode data.

2. Calculate the difference between the maximum time and the minimum time in the list-mode data: $t_d = t_{max} - t_{min}$

3. Divide the difference by the number of bins you have selected. Therefore, the size of each bin is $s = t_d / b$, where time is measured in microseconds.

4. Therefore, each bin is defined as follows: $(t_1 + (i - 1) \times s, t_1 + i \times s]$, where $i \in \{2, \ldots, b\}$. For the first bin, $i = 1$, the above is slightly modified in order to include the minimum time, t_1, within that bin. Therefore, the first bin is defined as follows: $[t_1, t_1 + s]$.

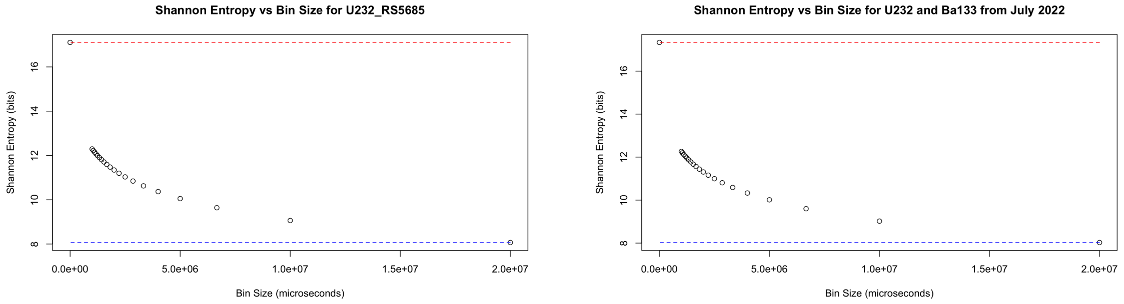
The above binning would result in bins that are equal in terms of the time they represent, but each bin was not required to have the same number of energy observations. Therefore, the number of events in each bin was dependant solely on the dataset being evaluated. Using different-sized bins of time would have led to an artificial effect on the Shannon entropy values because the length of time represented by the bin is correlated to the number of events included in the bin and therefore has an effect on the distribution of energy and time values within the bin. However, such variation in the length of time represented by each bin would not be accounted for by equation (5.6). It also would have been problematic to choose to have the same number of energy observations within each bin because it would have obscured the effects of the different datasets on the Shannon entropy. Doing this binning would have been the equivalent of comparing different-sized bins across datasets, which as stated above is problematic because such variation is not accounted for by equation (5.6).

Adjusting data based on bins

Following the creation of the bins, the list-mode data was adjusted. All values for times that fell within a certain bin were replaced with the maximum value that could have been obtained for said bin, $t_1 + i \times s$. In this way, each bin was its own spectrum because it contained the histogram of counts for each energy channel represented within said bin. Once this process was completed, the Shannon entropy for the new data was calculated by using equations (5.6) and (5.7).

5.3.4. *Results from Binning Experiments: Shannon Entropy*

In this section we demonstrate the effects of binning the list-mode data in time on Shannon entropy. For each source we used 20-second samples of data after removing the first 100 seconds of list-mode data. The sample of list-mode data was then binned in the way described in Section5.3.3. A few examples of the kind of plots generated to show how binning affected the Shannon entropy are displayed in Figure 5-3. All other plots for this analysis can be found in Appendix A.2. For each plot we chose to let $b \in \{1, \ldots, 20, n\}$. The red dashed line in those plots represents the maximum Shannon entropy obtained by our method of binning for a particular source's list-mode data ($b = n$). Similarly, the blue dashed line represents the minimum Shannon entropy obtained by our method of binning a particular source's list-mode data ($b = 1$). As was expected, the plots all show that binning in time results in a decrease in Shannon entropy.



(a) Shannon Entropy for Binning of U232 on July 2022

(b) Shannon Entropy for Binning of U232 & Ba133 on July 2022

Figure 5-3. Figures 5-3a and 5-3b display the Shannon entropy associated with the binning of data from the U232 and the combination of U232 and Ba133 datasets, respectively. The list-mode data associated with these plots was collected in July 2022.

Although these Shannon entropy calculations show that there is a loss of information when list-mode data is binned in time, this observation does not answer the question of whether or not the information that was lost is useful for radioisotope identification. Though such a question is difficult to answer, we attempted to provide evidence that would indicate there is a relationship between the temporal and energy information that is unique to the list-mode data that was collected. We made this attempt by comparing the Shannon entropy obtained when the list-mode data was binned to that obtained when binning randomly generated data. For the original list-mode data the order in which the energies appear is related to the order of times. The random list-mode data was created by shuffling the energies, which is intended to remove any potential relationship between the order of the energies and the time. For each source 100 random list-mode datasets were created. The absolute percent difference between the Shannon entropy results for the original list-mode data and each of the 100 random list-mode datasets was computed for each source as shown in equation (5.9).

$$\text{Absolute Percent Difference} = 100 \times \frac{|H(T,E)_r - H(T,E)_o|}{H(T,E)_r} \tag{5.9}$$

In equation (5.9), $H(T,E)_r$ represents the Shannon entropy associated with the random list-mode dataset after binning, and $H(T,E)_o$ represents the Shannon entropy associated with the original list-mode dataset after binning. The maximum difference among all 100 differences was obtained for each source. The results can be seen in Table 5-2 under the Maximum Absolute Difference column. The differences were also averaged over all 100 random list-mode datasets for each source in order to obtain the Mean Absolute Difference found in Table 5-2. The last column of Table 5-2 shows the standard deviation of the absolute percent difference. These results seem to indicate that the difference between the Shannon entropy associated with the original list-mode data and that associated with the randomly generated data is negligible, which implies that as far as the Shannon entropy metric is concerned it is difficult to differentiate between the list-mode data that was collected and that associated with a random process. It also implies that the decrease in Shannon entropy as the binning process is executed is not a unique feature of the list-mode data that was collected. Despite our inability to show that there are unique features in the list-mode data that result in a decrease in Shannon entropy as one bins in time, it is not possible for us to say that the time domain provides absolutely no benefit in the identification of radioisotopes.

Comparison of Shannon Entropy To Random Case			
Datasets: July 2022	Maximum Absolute Difference (%)	Mean Absolute Difference (%)	Standard Deviation (%)
Background	0.0884	0.0204	0.0179
Ba133	0.0519	0.0105	0.0093
Cs137	0.0747	0.0145	0.0127
U232	0.0339	0.0055	0.0051
U232 & Ba133	0.0232	0.0039	0.0036
Y88	0.0919	0.0187	0.0170
Datasets: April 2024			
Background	0.1319	0.0255	0.0225
Ba133	0.0825	0.0177	0.0146
U232	0.0974	0.0210	0.0172
U232 & Ba133	0.0513	0.0111	0.0095

Table 5-2. The table displays the metrics that were calculated from comparisons between the Shannon entropy of binning the source list-mode data and that of binning randomly generated dataset

5.3.5. Results from Binning Experiments: Fisher Information

Unbinned Fisher Information for λ_k

In the unbinned log-likehood function scenario, the Fisher information associated with the rate parameter for the background, λ_0, is calculated in the following way:

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta}|\mathcal{D}, K, B)}{\partial \lambda_0} &= \frac{\partial \left[\sum_{i=1}^{n} \log\left(\lambda_0 p_{E_0}(e_i)\right) - x_i \lambda_0\right]}{\partial \lambda_0}, \\
&= \sum_{i=1}^{n} \frac{\partial \left[\log\left(\lambda_0 p_{E_0}(e_i)\right) - x_i \lambda_0\right]}{\partial \lambda_0}, \\
&= \sum_{i=1}^{n} \left[\frac{\partial \log\left(\lambda_0 p_{E_0}(e_i)\right)}{\partial \lambda_0} - \frac{\partial x_i \lambda_0}{\partial \lambda_0}\right], \\
&= \sum_{i=1}^{n} \left[\frac{p_{E_0}(e_i)}{p_{E_0}(e_i)\lambda_0} - x_i\right], \\
&= \sum_{i=1}^{n} \left[\frac{1}{\lambda_0} - x_i\right]
\end{aligned}$$

$$-\mathbb{E}\left[\frac{\partial^2 \ell(\boldsymbol{\theta}|\mathcal{D}, K, B)}{\partial \lambda_0^2}\right] = -\mathbb{E}\left[\frac{\partial \sum_{i=1}^{n}\left[\frac{1}{\lambda_0} - X_i\right]}{\partial \lambda_0}\right], \tag{5.10}$$

$$= \sum_{i=1}^{n} \mathbb{E}\left[\frac{1}{\lambda_0^2}\right], \tag{5.11}$$

$$= \frac{n}{\lambda_0^2}, \tag{5.12}$$

where for background it is implied that $\pi_{k_1} = \pi_0 = 1$, $\lambda_{k_2} = \lambda_0$, and $p_{E_{k_1}} = p_{E_0}$

Similarly, the Fisher information associated with the rate parameter for the non-background sources, λ_1, is calculated as follows:

$$
\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta}|\mathcal{D},K,B)}{\partial\lambda_1} &= \frac{\partial\left[\sum_{i=1}^{n} \log\left(\sum_{k_2=0}^{1}\sum_{k_1=0}^{1} \pi_{k_1}\lambda_{k_2}p_{E_{k_1}}(e_i)\right) - \sum_{k_2=0}^{1} x_i\lambda_{k_2}\right]}{\partial\lambda_1},\\
&= \sum_{i=1}^{n} \frac{\partial\left[\log\left((\lambda_0+\lambda_1)\left(\pi_0 p_{E_0}(e_i)+\pi_1 p_{E_1}(e_i)\right)\right) - x_i\left(\lambda_0+\lambda_1\right)\right]}{\partial\lambda_1},\\
&= \sum_{i=1}^{n} \frac{\partial\left[\log\left(\lambda_0+\lambda_1\right) + \log\left(\pi_0 p_{E_0}(e_i)+\pi_1 p_{E_1}(e_i)\right) - x_i\left(\lambda_0+\lambda_1\right)\right]}{\partial\lambda_1},\\
&= \sum_{i=1}^{n} \left[\frac{\partial\log\left(\lambda_0+\lambda_1\right)}{\partial\lambda_1} + \frac{\partial\log\left(\pi_0 p_{E_0}(e_i)+\pi_1 p_{E_1}(e_i)\right)}{\partial\lambda_1} - \frac{\partial x_i\left(\lambda_0+\lambda_1\right)}{\partial\lambda_1}\right],\\
&= \sum_{i=1}^{n} \left[\frac{1}{\lambda_0+\lambda_1} - x_i\right]
\end{aligned}
$$

$$
\begin{aligned}
-\mathbb{E}\left[\frac{\partial^2\ell(\boldsymbol{\theta}|\mathcal{D},K,B)}{\partial\lambda_1^2}\right] &= -\mathbb{E}\left[\frac{\partial\sum_{i=1}^{n}\left[\frac{1}{\lambda_0+\lambda_1} - X_i\right]}{\partial\lambda_1}\right], && (5.13)\\
&= \sum_{i=1}^{n}\mathbb{E}\left[\frac{1}{\left(\lambda_0+\lambda_1\right)^2}\right], && (5.14)\\
&= \frac{n}{\left(\lambda_0+\lambda_1\right)^2}. && (5.15)
\end{aligned}
$$

Equations 5.12 and 5.15 show that for the unbinned log-likelihood function the Fisher information associated with the rate parameters is not expected to change as the bin size changes. This behavior is not surprising because the unbinned log-likelihood function does not account for binning and the total number of recorded events does not change when binning is used.

Binned Fisher Information

In the binned log-likehood function scenario, the Fisher information associated with the rate parameter for the background, λ_{B_0}, is calculated in the following way:

$$
\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta}|\mathcal{D},K,B)}{\partial\lambda_{B_0}} &= \frac{\partial\left[\sum_{j=1}^{m}\left(n_j\log\left(\lambda_{B_0}\right) - \lambda_{B_0} - \log(n_j!) + \sum_{i=1}^{n_j}\log\left(p_{E_0}(e_i)\right)\right)\right]}{\partial\lambda_{B_0}},\\
&= \sum_{j=1}^{m}\left[\frac{\partial\left[\left(n_j\log\left(\lambda_{B_0}\right) - \lambda_{B_0} - \log(n_j!)\right)\right]}{\partial\lambda_{B_0}} + \sum_{i=1}^{n_j}\frac{\partial\left[\log\left(p_{E_0}(e_i)\right)\right]}{\partial\lambda_{B_0}}\right],\\
&= \sum_{j=1}^{m}\left[\frac{\partial\left(n_j\log\left(\lambda_{B_0}\right)\right)}{\partial\lambda_{B_0}} - \frac{\partial\left(\lambda_{B_0}\right)}{\partial\lambda_{B_0}} - \frac{\partial\log\left(n_j!\right)}{\partial\lambda_{B_0}} + \sum_{i=1}^{n_j}\left[\frac{\partial\log\left(p_{E_0}(e_i)\right)}{\partial\lambda_{B_0}}\right]\right],\\
&= \sum_{j=1}^{m}\left[\frac{n_j}{\lambda_{B_0}} - 1\right]
\end{aligned}
$$

$$
\begin{align}
-\mathbb{E}\left[\frac{\partial^2 \ell(\boldsymbol{\theta}|\mathcal{D},K,B)}{\partial\lambda_{B_0}^2}\right] &= -\mathbb{E}\left[\frac{\partial\left[\sum_{j=1}^{m}\left(\frac{n_j}{\lambda_{B_0}}-1\right)\right]}{\partial\lambda_{B_0}}\right], \tag{5.16}\\
&= \sum_{j=1}^{m}\mathbb{E}\left[\frac{n_j}{\lambda_{B_0}^2}\right], \tag{5.17}\\
&= \sum_{j=1}^{m}\frac{\lambda_{B_0}}{\lambda_{B_0}^2}, \tag{5.18}\\
&= \frac{m}{\lambda_{B_0}} \tag{5.19}
\end{align}
$$

Similarly, the Fisher information associated with the rate parameter for the non-background sources, λ_1, is calculated as follows:

$$
\begin{aligned}
\frac{\partial\ell(\boldsymbol{\theta}|\mathcal{D},K,B)}{\partial\lambda_1} &= \frac{\partial\left[\sum_{j=1}^{m}\left(n_j\log\left(\sum_{k=0}^{1}\lambda_k B_j\right)-\sum_{k=0}^{1}\lambda_k B_j-\log(n_j!)+\sum_{i=1}^{n_j}\log\left(\sum_{k=0}^{1}\pi_k p_{E_k}(e_i)\right)\right)\right]}{\partial\lambda_1},\\
&= \sum_{j=1}^{m}\frac{\partial\left[\left(n_j\log\left(\lambda_0 B_j+\lambda_1 B_j\right)-\lambda_0 B_j-\lambda_1 B_j-\log(n_j!)\right))\right]}{\partial\lambda_1}+\\
&\frac{\partial\left[\sum_{i=1}^{n_j}\log\left(\pi_0 p_{E_0}(e_i)+\pi_1 p_{E_1}(e_i)\right)\right]}{\partial\lambda_1},\\
&= \sum_{j=1}^{m}\left[\frac{\partial\left(n_j\log\left(\lambda_0 B_j+\lambda_1 B_j\right)\right)}{\partial\lambda_1}-\frac{\partial\lambda_0 B_j}{\partial\lambda_1}-\frac{\partial\lambda_1 B_j}{\partial\lambda_1}-\frac{\partial\log(n_j!)}{\partial\lambda_1}\right]+\\
&\sum_{i=1}^{n_j}\left[\frac{\partial\log\left(\pi_0 p_{E_0}(e_i)+\pi_1 p_{E_1}(e_i)\right)}{\partial\lambda_1}\right],\\
&= \sum_{j=1}^{m}\left[\frac{n_j B_j}{\lambda_0 B_j+\lambda_1 B_j}-B_j\right],\\
&= \sum_{j=1}^{m}\left[\frac{n_j}{\lambda_0+\lambda_1}-B_j\right]
\end{aligned}
$$

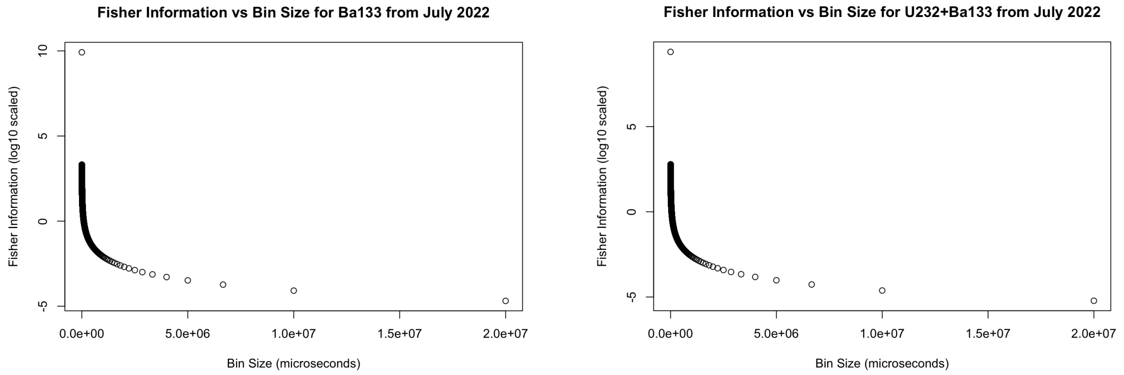
$$-\mathbb{E}\left[\frac{\partial^2 \ell(\boldsymbol{\theta}|\mathcal{D}, K, B)}{\partial \lambda_{B_1}^2}\right] = -\mathbb{E}\left[\frac{\partial\left[\sum_{j=1}^{m}\left(\frac{n_j}{\lambda_{B_0}+\lambda_{B_1}} - 1\right)\right]}{\partial \lambda_{B_1}}\right], \tag{5.20}$$

$$= \sum_{j=1}^{m} \mathbb{E}\left[\frac{n_j}{\left(\lambda_{B_0} + \lambda_{B_1}\right)^2}\right], \tag{5.21}$$

$$= \sum_{j=1}^{m} \frac{\left(\lambda_{B_0} + \lambda_{B_1}\right)}{\left(\lambda_{B_0} + \lambda_{B_1}\right)^2}, \tag{5.22}$$

$$= \frac{m}{\left(\lambda_{B_0} + \lambda_{B_1}\right)} \tag{5.23}$$

Equations 5.19 and 5.23 show that as the number of bins (m) decreases, so does the Fisher information associated with each rate parameter. Figure 5-4 shows this behavior empirically for a few datasets, but the remainder of the plots can be found in Appendix A.3.



(a) Fisher information for the rate parameter of Ba133 on July 2022

(b) Fisher information for the rate parameter of U232&Ba133 on July 2022

Figure 5-4. Figures 5-4a, and 5-4b display the change in the Fisher information associated with the rate parameter as the size of each bin is increased. The list-mode data associated with these plots was collected in July 2022.

This page intentionally left blank.

6. PROBABILISTIC CLASSIFIER MODEL

In this section, we introduce a probabilistic framework for RIID based on the well-established Poisson process representation of radioactive decay. The detection of radioactive isotopes that we are interested in involves collecting list-mode data, recording individual decay events characterized by both time stamps and energy levels. Given this data and random nature of radioactive decay, the Poisson process provides a natural mathematical model to describe the temporal behavior of such events.

Specifically, a Poisson process is a stochastic process that models the occurrence of events happening independently and at a constant average rate. This definition aligns with the underlying physics of radioactive decay, where the number of decays in a given time interval follows a Poisson distribution, and the time between decay events follows an exponential distribution. In the context of RIID, where multiple isotopes may be present along with background radiation, we can model the total number of observed decay events as a superposition of independent Poisson processes, each corresponding to a different isotope or background source.

By leveraging this Poisson process framework, we formulate a probabilistic classifier for RIID that aims to identify the most likely isotope(s) present in the observed data. This classifier is based on the observed interarrival times of events and energy levels, which are unique signatures for different isotopes. Using maximum likelihood estimation (MLE) techniques, we can infer the parameters governing the rate of decay and the energy distributions for each isotope. Given a new sequence of decay events, the classifier computes the likelihood of the observed data under each possible isotope hypothesis and selects the isotope with the highest likelihood as the predicted source of radiation.

6.1. Background: Poisson Processes

In this section, we first provide a formal description of a Poisson process, before explaining how it served as the foundation for our probabilistic classification models for RIID.

Definition 1. *A time-homogeneous Poisson process* $\{N_t\}_{t\geq 0}$ *with rate* $\lambda > 0$ *is a stochastic process with values in* $\mathbb{N}_{\geq 0}$ *satisfying:*

1. $\mathbb{P}(N_0 = 0) = 1$.
2. *The independent increment property: Given any choice* $n \in \mathbb{N}$ *and* $0 \leq t_0 < t_1 < t_2 < \cdots < t_n$*, the random variables* $N_{t_0}, N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, N_{t_3} - N_{t_2}, \ldots, N_{t_n} - N_{t_{n-1}}$ *are independent.*

3. *For any* $0 \le s < t$, $n \in \mathbb{N}_0$

$$\mathbb{P}(N_t - N_s = n) = \frac{(\lambda(t-s))^n e^{-\lambda(t-s)}}{n!}.$$

The number of events in $[s,t]$, $N_t - N_s$ *is a Poisson random variable of mean* $\lambda(t-s)$, *i.e.,* $N_t - N_s \sim Poi(\lambda(t-s))$, *which can be shown to be equivalent in distribution to* N_{t-s}.

Markov Property

A stochastic process N_t taking values in some countable set $\mathcal{N}$ is *Markov* if and only if, for any sequence of $0 \le t_1 < t_2 < \cdots < t_n$ and any $s_0, s_1, \ldots, s_n \in \mathcal{N}$

$$\mathbb{P}(N_{t_n} = s_n | N_{t_{n-1}} = s_{n-1}, \ldots, N_{t_0} = s_0) = \mathbb{P}(N_{t_n} = s_n | N_{t_{n-1}} = s_{n-1}).$$

Due to the Poisson Process's independent increment property, it is easy to see that for any $s_0, \ldots, s_n \in \mathbb{N}_{\ge 0}$ and $0 \le t_1 < t_2 < \cdots < t_n$

$$\begin{aligned}\mathbb{P}(N_{t_n} = s_n | N_{t_{n-1}} = s_{n-1}, \ldots, N_{t_0} = s_0) &= \mathbb{P}(s_n - s_{n-1} \text{events in} (t_{n-1}, t_n]) \\ &= \mathbb{P}(N_{t_n} = s_n | N_{t_{n-1}} = s_{n-1}),\end{aligned}$$

so that the Poisson Process is a Markov Process.

Interarrival time distribution

Given a Poisson process $N_t \sim \text{Poi}(\lambda t)$, let $(X_1, \ldots, X_n) := X_{1:n}$ be the interarrival times for the first n events such that $N_{X_1} = 1, N_{X_1+X_2} = 2, \ldots N_{X_1+\ldots,X_n} = n$. Now consider, for $t > 0$,

$$\mathbb{P}(X_1 > t) = \mathbb{P}(\text{no events in} [0,t]) = \mathbb{P}(N_t = 0) = e^{-\lambda t}.$$

$P(X_1 > t)$, also known as the survival function of X_1 implies the following cumulative distribution function (CDF) for X_1:

$$\mathbb{P}(X_1 \le t) = 1 - \mathbb{P}(X_1 > t) = 1 - e^{-\lambda t},$$

which is recognized as the CDF for an exponential distribution with rate parameter λ, i.e., $X_1 \sim \text{Exp}(\lambda)$.

Given $X_1 = t_1$, we can consider

$$\begin{aligned}\mathbb{P}(X_2 > t | X_1 = t_1) &= \mathbb{P}(\text{no events in} (t_1, t_1 + t] | X_1 = t_1) \\ &= \mathbb{P}(N_{t_1+t} - N_{t_1} = 0 | X_1 = t_1) \\ &= \mathbb{P}(N_t = 0) \\ &= e^{-\lambda t},\end{aligned}$$

yielding $X_2 \sim \text{Exp}(\lambda)$. Repeating this derivation for all remaining $X_j \in X_{1:n}$, we see that the interarrival times of a Homogeneous Poisson Process (HPP) are exponentially distributed with rate parameter λ.

Superposition

A HPP $N_t \sim \text{Poi}(\lambda t)$ has moment generating function (mgf)
$\mathbb{E}(e^{sN_t}) = \sum_{n=0}^{\infty} \frac{e^{sn}e^{-\lambda t}(\lambda t)^n}{n!} = e^{-\lambda t - \lambda t e^s} \sum_{n=0}^{\infty} \frac{e^{-\lambda t e^s}(\lambda t e^s)^n}{n!} = \exp(-\lambda t(1-e^s))$.

For n independent HPPs $N_t^1 \sim \text{Poi}(\lambda_1 t), \ldots, N_t^n \sim \text{Poi}(\lambda_n t)$, the summed quantity $N_t = \sum_{i=1}^n N_t^i$ has the mgf

$$\begin{aligned} \mathbb{E}(e^{s\sum_{i=1}^n N_t^i}) &= \Pi_{i=1}^n \mathbb{E}(e^{sN_t^i}) \\ &= \Pi_{i=1}^n \exp(-\lambda_i t(1-e^s) \\ &= \exp\left(-t\sum_i^n \lambda_i(1-e^s)\right), \end{aligned}$$

which is recognized as the mgf of a Poisson distribution with rate $\lambda = \sum_i^n \lambda_i$. Hence,
$N_t = \sum_{i=1}^n N_t^i \sim \text{Poi}(t\sum_{i=1}^n N_t^i)$.

Inhomogeneity

For a time inhomogeneous Poisson processes $\{N_t\}_{t\geq 0}$ with integrable rate function $\lambda(t) > 0$, one can show that

$$\mathbb{P}(N_t - N_s = n) = \frac{\Lambda^n(s,t)e^{-\Lambda(s,t)}}{n!},$$

where $\Lambda(s,t) := \int_s^t \lambda(u)du$.

6.2. Modeling list-mode data

In RIID, we collect list-mode data that contains *events* or decays, as a tuple: a time stamp with a discretized energy measured by the detector. A gamma ray detector records the data $\{(E(T_i), T_i)\}_{i=1}^n$, where the T_i denotes the time-stamp of the ith event (decay) and is such that $T_i > T_j$ for $j < i$. Given $T_i = t$, the random variable $E(t)$ measures the (discrete) energy recorded by the detector at time t and takes values in $\mathbb{N}$. However, the true energies are continuous in nature.

In an arbitrary environment with $K \geq 0$ isotopes and background, all $K+1$ sources are assumed to host environments in which gamma rays can be emitted randomly, but independently between each source. We assume that the gamma-ray detector is able to pick up on decays emitted from any of the $K+1$ sources, yet is unable to differentiate between sources at the time of measurement collection. To formalize this assumption, let us denote $N_t^0 \sim \text{Poi}(\lambda_0 t)$ as the HPP denoting *background decay counts*, and $N_t^i \sim \text{Poi}(\lambda_i t)$ as the HPP denoting decay counts from the i^{th} $(i = 1, \ldots, K)$ isotope present.

Given current methods and literature for RIID, particularly that standard temporal models of radioisotope decay are continuous time Markov Processes with distinct half lives (or rates of decay) per distinct isotope, each of the $K + 1$ sources is assumed to have a unique signature in both energy and time. Conversely, having a unique signature in time implies that the Poisson Process generating the decay counts for each of the $K + 1$ sources has a distinct rate, i.e., $\lambda_0 \neq \lambda_1 \neq \cdots \neq \lambda_K$.

As such, we model event (decay) counts recorded by the detector over time via the process $N_t = \sum_{i=0}^{K} N_t^i$, where $N_t^i \sim \mathrm{Poi}(\lambda_i)$. By the HPP superposition property, we have that any recorded event (decay) follows

$$N_t \sim \mathrm{Poi}\left(\sum_{i=0}^{K} \lambda_i t\right). \tag{6.1}$$

Further, the interarrival times $X_1, \dots X_n$ of decays counted by N_t follow $X_i \sim \mathrm{Exp}(\sum_{k=0}^{K} \lambda_k)$. The interarrival times are related to the time stamps of the first n events $T_0, \dots, T_n$ (with $T_0 = 0$) via

$$X_i = T_i - \sum_{j<i} T_j.$$

6.2.1. Parametric modeling

To model the energies E, we first note that the Breit-Wigner probability distribution [16] is often used to model the energy spectrum of gamma-rays emitted from excited nuclear states. Specifically, this distribution is a continuous probability distribution that describes the density of spectral lines arising from energy-level transitions in atomic nuclei. In this context, the following parameters are utilized to specify this *parametric* distribution:

- Peak Energy E_0: The peak energy representing the median energy value of the spectral line, corresponding to the energy of the transition responsible for the gamma ray emission.
- Full Width at Half Maximum (Γ): This quantity represents the width of the spectral line at half of its maximum height, characterizing the spread of energies around the median energy.

With E_0, Γ, the (continuous) probability of the gamma-ray detector measuring an energy e from an isotope follows:

$$p_E(e) = \frac{\Gamma/2\pi}{(e - E_0)^2 + (\Gamma/2)^2}, \tag{6.2}$$

which can be seen as a Cauchy distribution

$$p_E(E) = \frac{1}{\pi\left(\frac{\Gamma}{2}\right)\left[1 + \left(\frac{E-E_0}{\Gamma/2}\right)^2\right]}, \qquad E \in \mathbb{R}$$

with location parameter E_0 and scale parameter $\Gamma/2$. In practice since $E \geq 0$, the range can be restricted provided the normalization constant is updated.

To model *discrete* energies as they are collected from the detector, one can either use the density in Equation 6.2 and compute the cumulative distribution between the discretized intervals to obtain the probabilities of falling within each discretized bin or utilize a discrete version of the Cauchy distribution [17]:

$$p_E(e) = \frac{\tanh(\gamma\pi)}{\pi}\left(\frac{\gamma}{\gamma^2 + (e - E_0)^2}\right), \quad e \in \mathbb{Z}, \quad \gamma > 0,$$

by setting $\gamma = \Gamma/2$.

6.2.2. Nonparametric modeling

While there are a substantial number of methods that exist for modeling energies with parametric probability distributions, non-parametric methods also exist. Despite not having "closed form" representations, non-parametric models are flexible enough to capture underlying nuances in the data and are often much quicker to fit than their parametric counterparts.

One of the most famous non-parametric models is the histogram. A histogram estimates the probability density function (PDF) of a random variable by dividing the data range into a set of bins, where the height of each bin is proportional to the number of data points that fall within the bin. Mathematically, given a sample of data points $\{x_1, x_2, \ldots, x_n\}$, the histogram estimator for the probability density function $\hat{f}(x)$ is defined as:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}\left(\frac{x_i - x}{h}\right),$$

where: n is the total number of data points, h is the width of the bins (often called the bandwidth), $\mathbb{1}$ and is the indicator function in which $\mathbb{1}(A) = 1$ if A is true, and is zero otherwise.

The flexibility of the histogram lies in its ability to represent arbitrary shapes of the underlying distribution without assuming a specific parametric form. However, this flexibility comes at the cost of sensitivity to the choice of bin width h, which controls the smoothness of the estimate. A smaller h leads to a more detailed, but potentially noisier, estimate, while a larger h produces a smoother, but possibly less accurate, representation of the underlying distribution.

Beyond the simple histogram, more advanced non-parametric density estimation methods exist, such as kernel density estimation (KDE) [18], which can provide smoother approximations by using continuous kernels instead of the discrete bins of a histogram.

Whether parametric or nonparameteric, we assume each of the K isotopes and background radiation present invokes a mixture distribution of energies from their aggregated energy spectra. Specifically, in the presence of K isotopes and background, the energy collected by the detector during a decay event follows

$$p_E(e) = \sum_{i=0}^{K} \pi_i p_{E_i}(e), \tag{6.3}$$

where we model

$$p_{E_0}(e) = \sum_{b=1}^{B} p_b \frac{e^{-\alpha_b}\alpha_b^e}{e!},$$

to denote a Poisson mixture model of B components representing B background peak energies, each with mean $\alpha_b, b \in \{1, \dots, B\}$, and where $\sum_{b=1}^{B} p_b = 1$. Further, for $i > 0$ $p_{E_i}(e; \Gamma_i, E_0^i)$ denotes the full width at half maximum and peak (median) energies of isotope $i \in \{1, \dots, K\}$. Finally, the π_is such that $\sum_{i=1}^{K} \pi_i = 1$, provide probabilities of each of the $K + 1$ sources randomly emitting a recorded decay event.

6.2.3. *Marked point process setup*

A marked point process $(\mathcal{E}, \mathcal{T}, \mathcal{P})$ denotes random counting measure $\mathcal{N}$ defined on $\mathbb{E} \times \mathbb{T}$, where:

- $\mathcal{T}$ is the underlying point process representing the locations of the points,
- $\mathcal{E}$ is the set of possible marks,
- $\mathcal{P}$ is the probability distribution of the marks,
- $\mathcal{N}(B \times A)$ denotes the number of points of $\mathcal{T}$ in $A \subset \mathbb{T}$ with marks in $B \subset \mathbb{E}$, and
- For any finite collection of pairwise disjoint sets $\{(B_i, A_i) \mid i \in \mathbb{N}\}$, the random variables $\mathcal{N}(B_i \times A_i)$ are independent.

We define the marked list-mode point process as $(\mathcal{T}, \mathcal{E}, p_E) \in \mathbb{R}_{\geq 0} \times \mathbb{Z}_{+}$, defined by the set of marks that have probability measure p_E, $\mathcal{E}$ attached to each event in $\mathbb{R}_{\geq 0}$. A marked point process $(\mathcal{T}, \mathcal{E}, p_E)$ on $\mathbb{R}_{\geq 0} \times \mathbb{Z}_{+}$ is characterized by its *conditional intensity function* (CIF) $\lambda^*(t, e)$, such that for each $t \in \mathbb{R}_{\geq 0}$ and $e \in \mathbb{Z}_{+}$, $\lambda^*(t, e) > 0$ represents the intensity at temporal location t with energy e. It satisfies

$$\lambda^*(t, e) = \lim_{\substack{\delta t \to 0 \\ \delta e \to 0}} \frac{\mathbb{E}[\mathcal{N}(t + dt, e + de) - \mathcal{N}(t, e) | \mathcal{H}_t]}{\delta t \delta e},$$

where $\mathcal{H}_t$ denotes the history of the process (i.e., all joint events that have occurred before time t). Specifically, the CIF specifies the expected number of points of $\mathcal{T}$ in an infinitesimal neighborhood of t with mark e given the past history of the process up to time t, $\mathcal{H}_t$.

To define the CIF of the list-mode marked process, we note that

$$\lambda^*(t, e) = p_E(E; t)\bar{\lambda}(t),$$

where $p_E(E; t)$ denotes the probability of observing mark energy e given time-stamp t, and $\bar{\lambda}(t)$ denotes the CIF of the temporal point process only.

6.2.4. Likelihood Functions

The likelihood function of parameters $\boldsymbol{\theta} = \{(\pi_i, \lambda_i, (\theta)_{p_{E_i}})\}_{i=0}^{K}$ or the list-mode marked point process $(\mathcal{T}, \mathcal{E}, p_E)$ observed over a finite time interval $[0, T]$ and yielding data $\mathcal{D} = \{(e_{t_i}, t_i)\}_{i=1}^{n}$ is

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}|\mathcal{D}, K, B) &= \prod_{i=1}^{n} \left(\sum_{k=0}^{K} \pi_k p_{E_k}(e_i) \right) \bar{\lambda}(t_i) \exp\left(-\int_{t_i}^{t_{i+1}} \bar{\lambda}(u)\, du \right) \\
&= \prod_{i=1}^{n} \left(\sum_{k_2=0}^{K} \sum_{k_1=0}^{K} \pi_{k_1} \lambda_{k_2} p_{E_{k_1}}(e_i) \right) \exp\left(-(t_i - t_{i-1}) \sum_{k_2=0}^{K} \lambda_{k_2} \right) && (6.4) \\
\ell(\boldsymbol{\theta}|\mathcal{D}, K, B) &= \sum_{i=1}^{n} \log \left(\sum_{k_2=0}^{K} \sum_{k_1=0}^{K} \pi_{k_1} \lambda_{k_2} p_{E_{k_1}}(e_i) \right) - (t_i - t_{i-1}) \sum_{k_2=0}^{K} \lambda_{k_2}, \\
&= \sum_{i=1}^{n} \log \left(\sum_{k_2=0}^{K} \sum_{k_1=0}^{K} \pi_{k_1} \lambda_{k_2} p_{E_{k_1}}(e_i) \right) - x_i \sum_{k_2=0}^{K} \lambda_{k_2}, && (6.5)
\end{aligned}
$$

with $t_0 = 0$ and x_i denoting the ith interarrival time. Equation 6.5 will be referred to as the unbinned log-likelihood function.

In this application, exact times of events are typically not known as the time domain itself is *binned*. When the interarrival times are not known exactly, i.e., instead of having knowledge of $\{t_i\}_{i=1}^{n}$, we have $\{(n_i, B_i)\}_{i=1}^{m}$, which provides the number of events per bin of length B_i. In this case, the likelihood function is based on the Poisson counts in each bin, which are independent due to the independent increment property of the Poisson distribution:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}|\mathcal{D}, K, B) &= \prod_{j=1}^{m} \left(\frac{(\lambda_B)^{n_j} e^{-\lambda_B}}{n_j!} \prod_{i=1}^{n_j} \left(\sum_{k=0}^{K} \pi_k p_{E_k}(e_i) \right) \right) && (6.6) \\
&= \prod_{j=1}^{m} \left(\frac{\left(\sum_{k=0}^{K} \lambda_{B_k} \right)^{n_j} e^{-\sum_{k=0}^{K} \lambda_{B_k}}}{n_j!} \prod_{i=1}^{n_j} \left(\sum_{k=0}^{K} \pi_k p_{E_k}(e_i) \right) \right) && (6.7) \\
\ell(\boldsymbol{\theta}|\mathcal{D}, K, B) &= \sum_{j=1}^{m} \left(n_j \log \left(\sum_{k=0}^{K} \lambda_{B_k} \right) - \sum_{k=0}^{K} \lambda_{B_k} - \log(n_j!) + \sum_{i=1}^{n_j} \log \left(\sum_{k=0}^{K} \pi_k p_{E_k}(e_i) \right) \right), && (6.8)
\end{aligned}
$$

Here, m is the number of bins, n_j is the count of events in the j-th bin, B is the length of the each bin, e_i represents the i-th energy observation within the j-th bin, and $\lambda_{B_k} = \lambda_k B$. Equation 6.8 will be referred to as the binned log-likelihood function.

6.3. Probabilistic Classification

In this section, we detail how the binned and unbinned likelihoods can be used to facilitate our novel probabilistic classifier of isotopes from list-mode data.

6.3.1. Dataset and Set Definitions

In the following, we assume that for K isotopes, we have: 1) data from the background only, i.e., $\mathcal{D}_0 = \{(e^0_{t_i}, t^0_i)\}_{i=1}^{n^0_T}$, where $n^0_T = |\mathcal{D}_0|$ denotes the number of observations from the background over time period T, and 2) data from the same background obtained *with* isotope k as $\mathcal{D}_k = \{(e^k_{t_i}, t^k_i)\}_{i=1}^{n^k_T}$, over the same time period T. In general, we define a dataset containing background and some combination of the K isotopes $A \subseteq \{1, \ldots, K\}$ as $\mathcal{D}_A$. Due to the general observation process being a superposition of background and isotopes' (in A) processes, it is highly challenging to separately classify each observation as coming from either background or specific isotope from a single dataset $\mathcal{D}_A$. In the following, we see that to estimate the parameters $\boldsymbol{\theta}_k$ for isotope $k \in A$, a hierarchical set of datasets $\mathcal{D}_0$, $\mathcal{D}_A$, $\mathcal{D}_{A'}$, ... (where $A' \subset A$ with $|A'| = K - 2, A \subset \{1, \ldots, K\}$ with $|A| = K - 1$, and so forth, are required to be collected. For each non-empty subset $A' \subset A$ with $|A'| = |A| - 1$, we require a dataset $\mathcal{D}_{A'}$ that provides data for background and the isotopes in A'. Additionally, for each A', we require at least one dataset $\mathcal{D}_{A''}$ for every non-empty subset $A'' \subset A'$. This hierarchical data requirement ensures that we can accurately estimate the parameters for each isotope in A, by disentangling the contributions of each isotope and the background to the observed counts.

6.3.1.1. Energy Distribution Estimation

For a general dataset $\mathcal{D}$, the joint log-likelihood of energy counts and timestamps follows:

$$\ell(\boldsymbol{\theta}|\mathcal{D}, K) = \sum_{i=1}^{n} \log\left(\sum_{k_2=0}^{K} \sum_{k_1=0}^{K} \pi_{k_1}\lambda_{k_2} p_{E_{k_1}}(e_i)\right) - x_i \sum_{k=0}^{K} \lambda_k,$$

where $x_i := t_i - t_{i-1}$ and $t_{-1} := 0$. Since the energies and timestamps are modeled as independent random variables, the above log-likelihood can be maximized separately to obtain estimates of the parameters regulating the temporal and energy distributions.

6.3.1.2. Background Spectra Subtraction and Renormalization

To accurately classify isotopes in the presence of background radiation, we employ a method of spectral background subtraction and renormalization. Specifically, while the probability distribution of energies is a mixture:

$$p_E(e) = \sum_{i=0}^{K} \pi_i p_{E_i}(e)$$

that includes background, the raw *observed* energy counts over time period T are *unnormalized*. As such, if n_T observations $\{e_i\}_{i=1}^{n_T}$ have been collected by time T, the unnormalized (raw) distribution follows:

$$p_E^*(e) = \sum_{i=1}^{n_T} \mathbb{1}(e_i = e),$$

where $\mathbb{1}(e_i = e) = 1$ if observation $e_i = e$ and is zero otherwise. For continuous energy values, we elect to compute the unnormalized density of energies via the nonparametric histogram density approximator:

$$p_E^*(e) = \sum_{i=1}^{n_T} H\left(\frac{e - e_i}{h}\right),$$

where H denotes the kernel and h the bandwidth or smoothing parameter. Typical kernels include uniform, triangle, and Gaussian, and can be chosen such that the domain of the distribution satisfies constraints of the parameter, i.e., that $E > 0$. We choose this nonparametric method over other parametric models, discussed earlier, to reduce the number of parameters, and thus reliance, of the energies in studying the inclusion of temporal data on RIID.

We assume that pure background counts can be collected over time interval T to determine $p_{E_0}^*(e)$. With this assumption, when a single isotope is present (i.e., $K = 1$), measurements over the same time interval T can be collected to determine $p_{E|K=1}^*(e)$, which include counts from both the isotope and background. When obtaining data over the *same* time period T, therefore, we have:

$$p_{E_1}^*(e) = p_{E|K=1}^*(e) - p_{E_0}^*(e).$$

Assuming dataset collection over time intervals for multiple isotopes and background, this idea can then be extended to K isotopes by iteratively subtracting the background and already identified isotopic contributions. Specifically, for each isotope j in the presence of K isotopes, the unnormalized probability can be expressed as:

$$p_{E_j}^*(e) = p_{E|K=k}^*(e) - \sum_{\substack{i=0 \\ i \neq j}}^{k} p_{E_i}^*(e),$$

where $p_{E|K=k}^*(e)$ is the observed energy distribution for the mixture of background and K isotopes over the time interval T, and $p_{E_i}^*(e)$ are the previously identified energy distributions for background and other isotopes. This approach generalizes the subtraction and renormalization process to arbitrary K isotopes.

For a finite number of observations n_T from $p_{E|K=1}^*(e)$, we can ensure that $p_{E_1}^*(e) \geq 0$ by setting any negative values to zero. However, as $n_T \to \infty$, this practice becomes redundant.

More generally, for $K = k$ isotopes and $0 < j \leq k$, we can write:

$$p^*_{E_j}(e) = p^*_{E|K=k}(e) - \sum_{\substack{i=0 \\ i\neq j}}^{k} p^*_{E_i}(e). \tag{6.9}$$

Once computed, normalization can be pursued to obtain the estimated probability densities:

$$\hat{p}_{E_0}(e) = \frac{p^*_{E_0}(e)}{\sum_{e\in\mathcal{E}} p^*_{E_0}(e)}, \qquad \hat{p}_{E_j}(e) = \frac{p^*_{E_j}(e)}{\sum_{e\in\mathcal{E}} p^*_{E_j}(e)}, \qquad \hat{p}_{E|K=k}(e) = \frac{p^*_{E|K=k}(e)}{\sum_{e\in\mathcal{E}} p^*_{E|K=k}(e)},$$

where $j > 0$, and with the notion that the sums become integrals if the obtained energies are continuous.

The exercise of background spectra subtraction and renormalization is important for two reasons. First, it is necessary to remove background noise (counts that do not come from an isotope required to be detected), to determine the raw energy spectra of an isotope that could then be matched, and therefore detected, in future experiments. Second, in order to classify if a sequence of counts contains a specific isotope, the mixing proportions $\{\pi_i\}_{i=0}^K$ in the mixture density (given in Equation 6.3) are required to be estimated.

In particular, with the above methodology, the mixing proportions can be directly estimated by minimizing the difference between the normalized mixture density and the weighted sum of the background and isotope densities. Denoting:

$$\boldsymbol{\pi} = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_K \end{pmatrix},$$

as the vector of isotope mixing proportions with $\pi_0 = 1 - \sum_{j=1}^K \pi_j$, the least squares estimate $\hat{\boldsymbol{\pi}}$ of $\boldsymbol{\pi}$ can be computed via:

$$\hat{\boldsymbol{\pi}} = \arg\min_{\boldsymbol{\pi}} \left(\hat{p}_{E|K=k}(e) - \sum_{i=0}^{K} \pi_i \hat{p}_{E_i}(e) \right)^2 .$$

6.3.1.3. Temporal Process Estimation

Specifically, by defining the temporal process of counts in dataset $\mathcal{D}_A$ as N_t^A, we have from Equation 6.1 that:

$$N_t^A \sim \text{Poi}\left(\left[\lambda_0 + \sum_{k\in A} \lambda_k \right] t \right).$$

By further defining $\mathbf{X}_A = \{\mathbf{X}_A^n\}_{n=1}^{N_t^A}$ as the random interarrival sequence from dataset $\mathcal{D}_A$ and $\mathbf{x}_A = \{x_A^n\}_{n=0}^{|\mathcal{D}_A|}$ its realization, we have that:

$$X_A^j \sim \text{Exp}\left(\lambda_0 + \sum_{k\in A} \lambda_k \right).$$

With access to the correct datasets, therefore, we can estimate λ_0 and λ_i for each $i \in A$ via exponential MLE recursions. For any $n \geq 0$, we define the set $A_n = \{i_1, i_2, \ldots, i_n\}$ for $i_1, \ldots, i_n \in A$, and $i_1 \neq \cdots \neq i_n$, with data $\mathbf{x}_n \in \mathcal{D}_{A_n}$. We can recursively estimate every temporal rate in A via:

$$\begin{aligned}
\hat{\lambda}_{A_n} &= \bar{\mathbf{x}}_{A_n}^{-1} && \mathbf{x}_n \in \mathcal{D}_{A_n}, \quad A_n = \{i_1, i_2, \ldots, i_n\},\ i_1, \ldots, i_n \in A,\ i_1 \neq \cdots \neq i_n, \\
\hat{\lambda}_{i_n} &= \hat{\lambda}_{A_n} - \left(\hat{\lambda}_0 + \sum_{j=1}^{n-1} \hat{\lambda}_{i_j} \right), && \\
\hat{\lambda}_0 &= \bar{\mathbf{x}}_0^{-1} && \mathbf{x}_0 \in \mathcal{D}_0,
\end{aligned}$$

where $\bar{\mathbf{x}}_A = \sum_{i=1}^{n_{T_A}} x_A^i / n_{T_A}$ denotes the sample mean of the interarrival times $\mathbf{x}_A$.

Putting all of the above steps, we can compute the log-likelihood per isotope k This recursive estimation of the energy distributions for each isotope in A, and of the mixing proportions concludes parameter estimation across different isotopes and datasets.

6.3.1.4. Classification of an Unknown Sequence

Once we have obtained maximum likelihood estimates (MLEs) for the temporal rates $\{\hat{\lambda}_i\}_{i=1}^K$ and their corresponding energy distributions $\hat{p}_{E_0}(e), \hat{p}_{E_1}(e), \ldots, \hat{p}_{E_K}(e)$ from training sequences obtained via the hierarchical structure of datasets, we can classify a new sequence of observations $\mathcal{D}_{\text{new}} = \left\{(e_i', x_i')\right\}_{i=1}^n$ as belonging to one of K isotopes or purely background.

Specifically, given the newly observed sequence of decay events, the likelihood of the sequence can be computed for each isotope $k \in \{1, 2, \ldots, K\}$, using the trained model parameters. The total log-likelihood for isotope $k \in \{1, .., K\}$ combines the contributions of both the energy and time distributions:

$$\ell_k(\mathcal{D}_{new} | \hat{\boldsymbol{\theta}}_k) = \sum_{i=1}^{n'} \log\left(p_{E'}(e_i')\right) - x_i'(\hat{\lambda}_k + \hat{\lambda}_0),$$

where $\hat{\boldsymbol{\theta}}_k = \{\hat{\lambda}_0, \hat{\lambda}_k, \hat{p}_{E_0}(e), \hat{p}_{E_k}(e)\}$ for all $e \in \mathbb{E}$, $p_{E'}(e_i') = \hat{\pi}_0 p_{E_0}(e_i') + \hat{\pi}_k p_{E_k}(e_i')$ with $\hat{\pi}_k = 1 - \hat{\pi}_0$, $x_i' := t_i' - t_{i-1}'$ and $t_{-1}' := 0$.

The sequence is then classified as originating from isotope k^*, where

$$k^* = \arg \max_{k \in \{1, \ldots, K\}} \ell_k(\mathcal{D}_{new} | \hat{\boldsymbol{\theta}}_k),$$

meaning that the model selects the isotope with the highest likelihood given the observed energy and time sequence. We note that this method is equivalent to normalizing the log-likelihoods for each class to obtain a class probability. This approach therefore ensures that the most probable isotope responsible for the decay events in the test dataset is identified, taking both the temporal patterns and the energy spectrum into account.

Trivially, one can therefore classify isotopes based on either sole energy *or* temporal information via

$$k_e^* = \arg \max_{k\in\{1,\ldots,K\}} \sum_{i=1}^{n'} \log\left(p_{E'}(e_i')\right)$$
$$k_t^* = \arg \max_{k\in\{1,\ldots,K\}} -\sum_{i=1}^{n'} x_i'(\hat{\lambda}_k + \hat{\lambda}_0),$$

respectively.

6.4. The Extended α Model: Incorporating Signal-to-Noise Ratio (SNR)

Up until now, we have assumed that the test datasets contain list-mode sequences of decay events acquired in the same manner as that of the training dataset. However, this characteristic is an idealistic assumption that is, in practice, violated when collecting data from any given area. This violation is due to the fact that for testing, the source location is generally unknown, and thus the event counts, and potentially energy distributions, are likely to differ.

In this section, we extend the Poisson process model introduced earlier by incorporating a random component α that accounts for the Signal-to-Noise Ratio (SNR). The SNR reflects changes in the environment, such as varying distances between the radioactive source and the detector, which affect the observed decay rates. To account for this variability, we introduce a scaling factor α that modifies the decay rates for each isotope to adjust for changes in SNR during test conditions.

6.4.1. Model Description with α

As before, let λ_i denote the decay rate for isotope i as estimated from the training data under fixed conditions, such as a constant distance from the detector. In real-world scenarios, the observed decay rate λ_i' may vary due to fluctuations in SNR. We model this variation by introducing the scaling factor α, which modifies the rate as follows:

$$\lambda_i' = \alpha\lambda_i,$$

where ($\alpha > 0$) is a random variable representing the SNR effect. Higher (α) values correspond to higher SNR (closer detector-source distance), while lower α values correspond to lower SNR (further detector-source distance).

6.4.1.1. Bayesian MAP Inference for (α)

Since α can be highly variable, a Bayesian approach is necessary to represent any prior information one may have of α (e.g. that $\alpha \approx 1$) and to account for these uncertainties in a probabilistically just manner. In this manner, instead of marginalizing over α, i.e., integrating the likelihood over all values of $alpha$, which can lead to overly averaged likelihoods sensitive to the choice of prior, we estimate α_{MAP} (the maximum a posteriori estimate) for each test sequence. To achieve this estimation, we place a prior distribution on α and maximize the posterior distribution for each test sequence to find α_{MAP}.

Priors on (α)

We use a prior distribution on α to capture the expected variability in SNR. Since $\alpha > 0$, common choices for the prior could include:

- **Log-Normal Prior**: Suitable for modeling multiplicative variations in α:

  $$\alpha \sim \text{Log-Normal}(\mu_\alpha, \sigma_\alpha^2),$$

  where μ_α and σ_α^2 are the mean and variance of $\log(\alpha)$. With this prior, known expected values of α, e.g. $\alpha \approx 1$ can be easily incorporated by setting the prior mean and variances accordingly.

- **Gamma Prior**: Useful for skewed distributions of α, where the SNR tends to cluster around certain values:

  $$\alpha \sim \text{Gamma}(k_\alpha, \theta_\alpha),$$

  where k_α is the shape parameter and θ_α is the rate parameter. Again, these hyperparameters can be set according to any prior knowledge one may have of the level of SNR.

Posterior Likelihood and α_{MAP}

Given the test data $\mathcal{D}_{\text{test}} = \{(e_i', t_i')\}_{i=1}^{n'}$, we compute the posterior distribution of α using the likelihood of the observed data under isotope k:

$$p(\alpha | \mathcal{D}_{\text{test}}, \boldsymbol{\theta}_k) \propto p(\{(e_i', t_i' | \boldsymbol{\theta}_k)\} | \alpha) \pi(\alpha).$$

Here, $\pi(\alpha)$ denotes the prior density of α, and
$p(\{(e_i', t_i' | \boldsymbol{\theta}_k) \propto (\pi_k p_{E_k}(e_i') + \pi_0 p_{E_0}(e_i'))(\lambda_0 + \alpha\lambda_k) \exp(-x_i'(\lambda_0 + \alpha\lambda_k))$, which is based on the modified decay rates. Subsequently, this posterior can be maximized to obtain α_{MAP}, i.e.,

$$\alpha_{\text{MAP}} = \arg\max_\alpha p(\alpha | \{(e_i', t_i')\}),$$

which is computed for each new test sequence under the previously obtained MLEs $\hat{\boldsymbol{\theta}}$ from the training data only.

6.4.1.2. Posterior Density and Classification

Once α_{MAP} is estimated, we can compute the posterior density for a new test sequence using α_{MAP}. This process transforms the log-likelihood of a new sequence $\mathcal{D}_{\text{new}} = \{(e'_i, t'_i)\}_{i=1}^{n'}\}$ under isotope k from:

$$\ell_k(\mathcal{D}_{\text{new}}|\hat{\boldsymbol{\theta}}_k) = \sum_{i=1}^{n'} \log\left(\hat{\pi}_k \alpha_{\mathrm{MAP}} \hat{p}_{E_k}(e'_i) + \hat{\pi}_0 \hat{p}_{E_0}(e'_i)\right) - x'_i(\alpha_{\mathrm{MAP}} \hat{\lambda}_k + \hat{\lambda}_0),$$

where $\hat{\boldsymbol{\theta}}_k = \{\hat{\lambda}_0, \hat{\lambda}_k, \hat{p}_{E_0}(e), \hat{p}_{E_k}(e)\}$, to the new form:

$$\ell_k(\mathcal{D}_{\text{new}}|\hat{\boldsymbol{\theta}}_{\alpha,k}) = \log \pi(\alpha_{\mathrm{MAP}}) + \sum_{i=1}^{n'} \log\left(\hat{\pi}_k \alpha \hat{p}_{E_k}(e'_i) + \hat{\pi}_0 \hat{p}_{E_0}(e'_i)\right) - x'_i(\alpha_{\mathrm{MAP}} \hat{\lambda}_k + \alpha_{\mathrm{MAP}} \hat{\lambda}_0),$$

where $\hat{\boldsymbol{\theta}}_{\alpha,k}$ now includes $\hat{\alpha}_{\mathrm{MAP}}$, and $\pi(\alpha_{\mathrm{MAP}})$ represents the prior evaluated at α_{MAP}.

The updated parameter set for isotope k is now:

$$\hat{\boldsymbol{\theta}}_{\alpha,k} = \{\hat{\alpha}_{\mathrm{MAP}}, \hat{\lambda}_0, \hat{\lambda}_k, \hat{p}_{E_0}(e), \hat{p}_{E_k}(e)\},$$

for all $e \in \mathbb{E}$.

6.4.1.3. Classification Procedure

For a new sequence in $\mathcal{D}_{\text{new}}$, the final classification is performed by computing the updated log-likelihood for each isotope k using the estimated α_{MAP} and selecting the isotope with the highest log-likelihood:

$$k^* = \arg \max_{k \in \{1,\ldots,K\}} \ell_k(\mathcal{D}_{\text{new}}|\hat{\boldsymbol{\theta}}_{\alpha,k}).$$

This approach allows us to incorporate the variability due to SNR while retaining a discriminative classification criterion based on the MAP estimate of α. By doing so, we avoid the averaging over α, which can lead to sensitivity to the prior, and instead focus on the most probable value of α given the observed data.

7. CLASSIFIER ANALYSIS

Using the original probabilistic classification models described in Chapter 6 and the different datasets described in Chapter 4, several experiments were conducted to determine whether temporal information in the form of interarrival times between events provides some benefit over solely using energy information for the classification of radiological source material. Interarrival time is defined as the length of time in microseconds that has passed between the previous event and the current event. In each experiment, the first 100 seconds of list-mode data from each dataset was ignored. This exclusion was done because the first 60 seconds of data generally represent collection setup time. In each experiment 100,000 consecutive events were kept for each source in the datasets relevant to that experiment. Seventy-percent (70,000) of the events from each source were placed in a training set. Each test set varied based on the experiment that was being conducted, but the size of each test set remained the same across experiments. Only 3,333 samples from each source were used in each test set. Each sample was comprised of 100 events. Additionally, the accuracy results that are displayed in Table 7-1 represent the accuracy of the model when the model's predicted class is the class with the maximum likelihood value.

Information about the different experiments and the results related to them are provided below. Each experiment represents the use of a different training and test set. We note here that the combination source U-232+Ba-133 was considered as a *single* isotope[1] so that individual isotope classification as described in Section 6.3 could be conducted.

Experiment 1: July 2022

In this experiment the July 2022 dataset was used for training and testing. As mentioned in Chapter 4 this dataset includes events collected from background, Y-88, Cs-137, U-232, Ba-133, and U-232+Ba-133. The test set for each source was derived from the 30,000 events that were not used in the training set. In this way, the test set and the training set come from the same data distribution.

Experiment 2: April 2024

In this experiment the April 2024 dataset was used for training and testing. As mentioned in Chapter 4 this dataset includes events collected from background, U-232, Ba-133, and U-232+Ba-133. The test set was derived in a similar way as to that described for Experiment 1. In this way, the test set and the training set come from the same data distribution.

---

[1] Classification of multiple isotopes in an individual sample is seen as an exciting extension of the probabilistic classifier we have developed, but is beyond the scope of this preliminary work that will be studied in the future.

Experiment 3: D1 to D1

In this experiment the D1 dataset collected in August 2024 was used for training and testing. As mentioned in Chapter 4 this dataset includes events collected from background, Y-88, Cs-137, Ba-133, and Cs-137+Ba-133. The test set was derived in a similar way as to that described for Experiment 1. In this way, the test set and the training set come from the same data distribution. This arrangement also means that the distances at which the sources are placed is remaining constant from the training set to the test set.

Experiment 4: D1 to D2

In this experiment the D1 dataset collected in August 2024 was used for the training set whereas the D2 dataset collected in August 2024 was used for the test set. As mentioned in Chapter 4 this dataset includes events collected from background, Y-88, Cs-137, Ba-133, and Cs-137+Ba-133. The test set for each source was derived from 30,000 events found in the D2 dataset for each source. Therefore, the test set and the training set come from two different data distributions. This arrangement also means that the distances at which the sources are placed is increasing from the training set to the test set.

Experiment 5: D2 to D1

In this experiment the D2 dataset collected in August 2024 was used for the training set whereas the D1 dataset collected in August 2024 was used for the test set. As mentioned in Chapter 4 this dataset includes events collected from background, Y-88, Cs-137, Ba-133, and Cs-137+Ba-133. The test set for each source was derived from 30000 events found in the D1 dataset for each source. Therefore, the test set and the training set come from two different data distributions. This arrangement also means that the distances at which the sources are placed is decreasing from the training set to the test set.

Experiment 6: D2 to D2

In this experiment the D2 dataset collected in August 2024 was used for training and testing. As mentioned in Chapter 4 this dataset includes events collected from background, Y-88, Cs-137, Ba-133, and Cs-137+Ba-133. The test set was derived in a similar way as to that described for Experiment 1. In this way, the test set and the training set come from the same data distribution. This arrangement also means that the distances at which the sources are placed is remaining constant from the training set to the test set.

Experiment 7: Mix to Mix

In this experiment the D1 and D2 datasets collected in August 2024 were combined and then used to create the training and test set. As mentioned in Chapter 4 this dataset includes events collected from background, Y-88, Cs-137, Ba-133, and Cs-137+Ba-133. The test set for each source was derived from 30,000 events found in the combined dataset for each source. Therefore, the test set and the training set come from a combination of two different data distributions. This arrangement also means that the training and test sets are representative of the two different distances at which events were collected for each source.

Accuracy Results			
Experiment:	Time Model	Energy Model	Energy & Time Model
1: July 2022	0.7107	0.5498	0.6565
1-α: July 2022	0.7018	0.5498	0.6200
2: April 2024	0.8774	0.3890	0.6898
3: D1 to D1	0.4850	0.7368	0.8249
3-α: D1 to D1	0.3334	0.5436	0.8479
4: D1 to D2	0.1787	0.3864	0.5307
4-α: D1 to D2	0.3333	0.4782	0.7504
5: D2 to D1	0.2509	0.6636	0.7404
5-α: D2 to D1	0.3333	0.8799	0.9936
6: D2 to D2	0.5101	0.5476	0.7451
6-α: D2 to D2	0.3333	0.8372	0.9699
7: Mix to Mix	0.5037	0.6249	0.7488
7-α: Mix to Mix	0.3333	0.6442	0.9201

Table 7-1. The table displays the accuracy results that were obtained for each experiment.

7.1. Discussion of Experimental Results

In Table 7-1, we present experimental results of all datasets when trained and tested on the initial probabilistic classifier developed, and the extended α model that incorporates variabilities in the SNRs between training and testing datasets.

These results show that regardless of the experiment that was conducted, the dual energy and time models always outperformed the energy models (i.e., the spectral analyses only) in terms of accuracy. However, the benefit that can be attributed to adding temporal information to the model varies based on the experiment that was conducted, mainly due to each isotope's distinct decay parameters. For example, some of the success of the time models of experiments 1 and 2 may be due to the models' reliance on the count rate as a way of differentiating classes. Experiments 3 and 6 are intended to reduce the model's ability to rely on count rates by ensuring that during the collection process each source is positioned at a distance from the detector that forces the count rates between different sources to be similar. The similarity between these count rates could explain why the time models for experiments 3 and 6 perform worse than those for experiments 1 and 2. Under experiments 4 and 5 there is an additional drop in performance, which may be attributed to the fact that the training and test sets have a different distribution from each other. Within each training and test set of experiments 4 and 5 the count rate was controlled in the same way as described for experiments 3 and 6. Additionally, D2 represents counts that were collected for sources held farther away from the detector than the sources in D1. This difference means that in experiment 4 the training set represents data that has a higher source to background ratio than the test set whereas in experiment 5 the opposite is true. Therefore, the difference in performance between the models in experiment 4 and experiment 5 may be attributed to the shift in the data distribution between the training and test sets. However, using the extended α model that

		Predicted Class	
		Class of Interest	Other Class
Actual Class	Class of Interest	a	b
	Other Class	c	d

Table 7-2. This table defines the values that are used to calculate sensitivity and specificity.

specifically integrates uncertainties in SNRs between training and testing sets results in a much more powerful classifier that overcomes the drawbacks of the initial classifier, as presented from the results of experiments 3–6. Additionally, implementation on the first experiment shows only a slight change in the estimated classes between either model, highlighting the potential power of the extended model in varied, and potentially highly complex, data collection scenarios.

7.2. Receiver Operating Characteristic Curves

Receiver operating characteristic (ROC) curves are a tool used for assessing the performance of a binary classifier, comparing the performance of binary classifiers, and establishing a threshold likelihood value at which a data point is said to belong to a specific class [19]. The ROC curve typically has sensitivity on the y-axis and specificity or the false positive rate (FPR) on the x-axis. Sensitivity is defined as the ratio of the number of observations that were correctly associated to the class of interest by the model to the total number of observations that belong to the class of interest. On the other hand, specificity is defined as the ratio of the number of observations that were correctly associated to the class that was not of interest by the model to the total number of observations that do not belong to the class of interest [20]. Using Table 7-2, sensitivity and specificity can be defined as shown in equations 7.1 and 7.2.

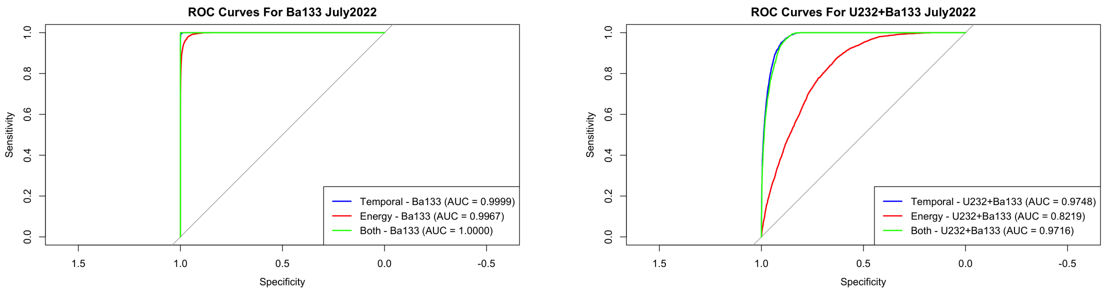
$$\text{Sensitivity} = \frac{a}{a+b} \tag{7.1}$$

$$\text{Specificity} = \frac{d}{c+d} \tag{7.2}$$

The letters shown in Table 7-2 can be defined as follows: a is the number of true positives, b is the number of false negatives, c is the number of false positives, and d is the number of true negatives.

The nonparametric ROC curve is then obtained by calculating the sensitivity and specificity for a given binary classifier at several threshold values. The threshold values are not displayed on the ROC curve, but higher threshold values mean that a larger likelihood value is needed in order for the model to classify an observation as the class of interest. This notion in turn means that higher threshold values usually result in a higher specificity and a lower sensitivity. Thus, sensitivity and specificity values associated with higher threshold values will be found closer to the lower left corner of the ROC curve, whereas the opposite is true for lower threshold values.

Often the area underneath the ROC curve (AUC) is used as a way to summarize the curve and as a way of comparing classifiers. The AUC, unlike the error rate or accuracy, is independent of the threshold value [21]. Let $\hat{p} = P(\hat{x}_i = 0|x_i = 1)$ be the estimated probability of incorrectly assigning a randomly chosen member of one class to another class and $\hat{q} = P(\hat{x}_i = 0|x_i = 0)$ be the estimated probability of correctly assigning a randomly chosen member of the other class, then the AUC can be defined as the probability that $\hat{p}$ is smaller than $\hat{q}$ [21]. Therefore, better performing models are associated with larger AUC values and AUC values range between 0 and 1 [20]. Examples of the ROC curves that were generated and their associated AUC values are shown in Figure 7-1. Additional ROC curves may be found in Appendix A.4. These AUCs and ROC curves were calculated by considering the class of interest as one class and putting all other classes into a second class. In this way, every ROC curve and AUC value only represents a two class scenario.



(a) Experiment 1: ROC for model performance on Ba133

(b) Experiment 1: ROC for model performance on U232 & Ba133

Figure 7-1. Figures 7-1a and 7-1b display the ROCs associated with the performance of the time, energy, and energy & time models on different sources. The list-mode data associated with these plots was collected in July 2022.

Since AUC is only applicable to situations where there are only two classes, additional analysis was conducted using the M measure presented in [21], which is the average of all pairwise AUCs and can be calculated as shown in equation 7.3. In equation 7.3, s represents the number of classes and AUC_{ij} represents the pairwise AUC between the i^{th} and j^{th} classes. It is important to note that these pairwise AUCs are different from the AUCs shown in Figure 7-1 and those found in Appendix A.4. The M measure is also referred to as the “multi-class AUC" [22].

$$M = \frac{\sum_{i \neq j} AUC_{ij}}{s(s-1)} \tag{7.3}$$

The results presented in Table 7-3 and the ROC curves and AUC values found in Appendix A.4 indicate that the energy & time model outperforms the energy model in every experiment that both 1) did not involve a change in the distribution of the data between the training and test set, and 2) did involve a change in distribution (see Figures A-20, A-21 and A-16 in Appendix A.4). These results highlight the flaw in the algorithm presented in Section 6.3, which exploits the proportion of background that is present in the signal for testing. When the proportion of background present in the signal changes between the training set and the test set, the extended α model, which accounts for the SNR (different distances), has seemed to overcome the flaws of the first model and provide accurate classification.

Multi-class AUC Results			
Experiment:	Time Model	Energy Model	Energy & Time Model
1: July 2022	0.9454	0.9105	0.9644
2: April 2024	0.9742	0.4845	0.9774
3: D1 to D1	0.7905	0.9614	0.9772
4: D1 to D2	0.5676	0.9171	0.9077
5: D2 to D1	0.5934	0.9167	0.9146
6: D2 to D2	0.8197	0.9313	0.9599
7: Mix to Mix	0.8049	0.9462	0.9662

Table 7-3. The multi-class AUC results computed for each experiment and model type.

7.3. McNemar's Test

The classifier results seem to indicate that a model that uses both interarrival time and energy channels in order to classify the sources outperforms a model that only uses energy channels. In order to determine whether the two classifiers are significantly different from each other, McNemar's Test is used. McNemar's test has been used and is frequently recommended as a nonparametric test for comparing the accuracy or error rate of two classifiers [23, 24].

		Classifier 2	
		Wrong	Correct
Classifier 1	Wrong	w	x
	Correct	y	z

Table 7-4. The matrix associated McNemar's Test.

In order to understand McNemar's test and how it is applied for comparing two classifiers, consider first the matrix in Table 7-4. In Table 7-4, w represents the number of samples that were incorrectly classified by both classifiers, x represents the number of samples that were correctly classified by classifier 2 and incorrectly classified by classifier 1, y represents the number of samples that were incorrectly classified by classifier 2 and correctly classified by classifier 1, and z represents the number of samples that were correctly classified by both classifiers. Therefore, if W, X, Y, and Z represent the random variables associated with the counts w, x, y, and z, and m is the total number of samples (i.e., $m = W + X + Y + Z$), then the error rates for classifier 1 and classifier 2 are $p_1 = \frac{W+X}{m}$ and $p_2 = \frac{W+Y}{m}$, respectively. Thus, a null hypothesis that states that both error rates are equal, is equivalent to a null hypothesis that states that $p_X = p_Y$. Let $n = X + Y$, then it can be shown that

$$X \sim Binomial\left(n, p = \frac{X}{X+Y}\right),$$

which means that X follows a binomial distribution with parameters n and p, where n represents the number of events in both X and Y and p represents the probability that an event is in category X [11]. Then, the null and alternative hypotheses associated with McNemar's test are as follows:

$$H_0 : p = \frac{1}{2} \quad H_1 : p > \frac{1}{2} \tag{7.4}$$

In hypothesis testing, there are two types of errors that are monitored: type 1 error and type 2 error. Type 1 error is defined as the error associated with rejecting the null hypothesis when the null hypothesis is true and type 2 error is defined as the failure to reject the null hypothesis when the alternative hypothesis is true. For the above hypothesis test, type 1 error can be defined as stating that the error rate for classifier 1 is larger than the error rate for classifier 2 when in fact both have the same error rate. On the other hand, type 2 error can be defined as failing to claim that the error rate for classifier 1 is greater than the error rate for classifier 2 when this inequality is actually true. Since our hypothesis is that including temporal information in our model is beneficial, we consider our energy-based model to be classifier 1 and the model that includes both energy and temporal information to be classifier 2. The next step in the hypothesis testing process was to set the maximum allowable type 1 error rate, which is also referred to as the level of significance, α. In order to stick to convention, we decided to set $\alpha = 0.05$ [11]. Table 7-5 shows the P-values that were obtained for each of the respective experiments that were conducted. In this case, a P-value can be interpreted as the probability of observing at least x samples that were incorrectly classified by the energy model and correctly classified by the model that includes energy and temporal information given that $p = \frac{1}{2}$. Therefore, under the hypothesis test in Equation 7.4, the smaller the P-value the more significant the difference between the two classifiers. This difference would indicate that the energy model performs significantly worse than the model that includes temporal and energy data. Equation 7.5 shows the equation that was used to calculate the P-values that are displayed in Table 7-5.

$$\text{P-value} = \left(\frac{1}{2}\right)^n \sum_{i=x}^{n} \binom{n}{i} \tag{7.5}$$

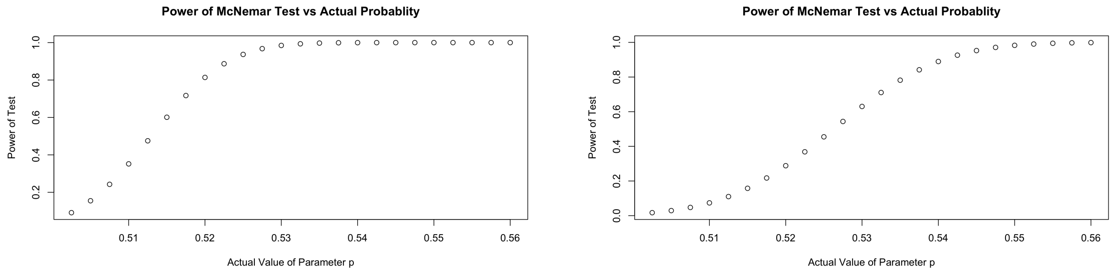
The results presented in Table 7-5 indicate that under each experiment the error rate for the energy model is significantly larger than the error rate for the energy & time model. These results seem to indicate that it is beneficial too include temporal information for the classification of radioactive material.

McNemar's Test Results	
Experiment:	P-value
1: July 2022	~ 0
2: April 2024	~ 0
3: D1 to D1	~ 0
4: D1 to D2	~ 0
5: D2 to D1	~ 0
6: D2 to D2	~ 0
7: Mix to Mix	~ 0

Table 7-5. Results of running McNemar's test to compare the error rate of the model that used both the energy and temporal information to the model that only used energy information for classification for each experiment.

Power Analysis

In order to ensure that the McNemar's tests were able to properly discriminate between the null and alternative hypotheses, we analyzed the power function for each test [11]. The plots of the power function show how the probability of rejecting the null hypothesis changes as the number of observations that were incorrectly labeled by the energy model and correctly labeled by the energy & time model increases. An increase in p represents an increase in the number of observations that were incorrectly labeled by the energy model and correctly labeled by the energy & time model. Figure 7-2 shows a couple examples where the McNemar's test is shown to be fairly sensitive to the change in p. These results were present in all plots of power as a function of p. The other power plots can be found in Appendix A.5. These results provide evidence to our claim that the tests were able to effectively discriminate between the null and alternative hypotheses.



(a) Experiment 2: Power At Several Values For p

(b) Experiment 5: Power At Several Values For p

Figure 7-2. Figures 7-2a and 7-2b display the power of the McNemar's test associated with the stated experiment as the value of the p parameter is increased.

8. CONCLUSION

In this work we asked two simple questions, what is the nature of the temporal domain information of list-mode gamma-ray data and is that temporal domain information beneficial to the application of radioisotope identification. We began with a hypothesis that Hidden Markov Models may be a beneficial method for utilizing temporal information in classifying radioactive sources given the close similarities to the properties of HMMs and the properties of radiation decay chain physics. However, we quickly realized through application and discussion with subject matter experts that this approach would be impractical, if not impossible, to implement. We confidently conclude that HMMs are not a good tool to employ for radiation data sets in the task of RIID.

We took a deep dive into understanding the nature of the time domain content of our list-mode datasets. We looked at several methods of discovering temporal correlations via Pearson's correlation, Spearman rho, autocorrelation, and crosscorelation methods. All these experiments concluded with finding no evidence of correlations between the time and/or energy space of our radiation datasets. We also investigated distributional analysis of our data sets, specifically looking at box plots to identify characteristics of the data sets that may lead to discernible discriminators between radioactive sources. These experiments indicated that the distributions of interarrival times are relatively stable across consistent collection windows and that these distributions do show meaningful differences across radioactive sources. Variations in the distributions for the same source did occur when collection windows increased or decreased.

Further, we looked at Shannon and Fisher information theory analysis of our datasets. Specifically we wanted to know if there was a measurable quantification of information loss when moving from list-mode data to spectral data. Our results indicate a positive affirmation of this hypothesis in both the definitions of Shannon Information and Fisher Information. This affirmation is a strong result to quantify as it is a precursor support to our ultimate question of the benefits of the temporal domain information. Knowing that there is a quantifiable amount of information loss, we were well prepared to investigate methods for exploiting this information in the application of RIID.

To provide supportive evidence to the usefulness of temporal information for RIID, we devised a probabilistic classification method that could use temporal information and energy information alone, or both combined. Given that the classifier is the same, and the nature of the data is the only thing changing, we establish an experiment that can determine if the inclusion of temporal information improves the performance of RIID. Our classifier is based on distributional estimates of the temporal and energy domain data.Through adequately defined testing and training datasets, we rigorously assessed the performance of our classifier under interarrival time data only, spectral data only, or both, through the use of ROC curves, McNemar tests, and analysis of the power function for each test. The results concluded that in all cases, it was always more performant to include temporal domain information when classifying. For any case in which it was not more performant to include temporal information, the performance differences between a spectral only classifier and the combination classifier was negligible.

The results of this study are promising and support future investigation into temporal information understanding and methods of utilizing temporal domain information for RIID applications. GADRAS[14] algorithms and modeling are the *defacto* standard for RIID today and has been in constant development and improvement for nearly 40 years. Our effort was a $100k funded effort spread out over 9 months. We do not, rightfully so, claim any comparison or improvements over GADRAS; we do, however, provide evidence and support to the potential benefit of including temporal domain information into RIID algorithms. We think we have shown sufficient evidence to warrant future investigation and hope that this research will lead to new methods, or extensions of current methods, augmented by temporal information that can improve RIID for the broader community of constituents.

9. FUTURE WORK

The body of work presented here was a first step in providing a rigorous understanding of the temporal domain information of radiological list-mode data collected from radiation sensors for applications of interest to the National Security Programs Investment Area. Our results provide a foundation for continued efforts in better leveraging the temporal domain information that we have identified. We feel we have provided a deeper understanding of the nature of list-mode radiation data and provided a foray into finding better temporal metrics and methods of leveraging temporal domain information for radioisotope identification applications. With that the following is a list of next steps that could be taken to continue this research forward (in no particular order):

1. Determining an effective sample size for the time domain of each dataset.
2. Extension of the probabilistic classifier to the task of source separation
3. Exploring other metrics besides interarrival time to develop discriminate features for use in a classifier
4. Applying the methods of this work to non-stationary sources
5. Investigating the use of event-based machine learning techniques, such as stable state-space modeling, for the analysis of list-mode data.
6. Apply real list-mode data to the N4PA spiking neural algorithm for RIID to determine if the event-based temporal algorithm performs similarly on list-mode data.
7. Extend the classifier framework to higher dimensional representations of the list-mode data as a means of better discriminating the temporal information content of the source.

This page intentionally left blank.

REFERENCES

[1] CJ Sullivan, ME Martinez, and SE Garner. Wavelet analysis of sodium iodide spectra. *IEEE Transactions on Nuclear Science*, 53(5):2916–2922, 2006.

[2] Kyle J Bilton, TH Joshi, MS Bandstra, JC Curtis, BJ Quiter, RJ Cooper, and K Vetter. Non-negative matrix factorization of gamma-ray spectra for background modeling, detection, and source identification. *IEEE Transactions on Nuclear Science*, 66(5):827–837, 2019.

[3] Eric Clarkson and Meredith Kupinski. Quantifying the loss of information from binning list-mode data. *JOSA A*, 37(3):450–457, 2020.

[4] Corinne Teeter, Aaron Hill, James Aimone, Craig Vineyard, Peter Helfer, and Dhireesha Kudithipudi. Arnie: Autonomous reconfigurable neural intelligence at the edge. Technical report SAND2023-14089, Sandia National Laboratories, Albuquerque, New Mexico 87185 and Livermore, California 94550, November 2023.

[5] Scott J Wilderman, Neal H Clinthorne, Jeffrey A Fessler, C-H Hua, and W Leslie Rogers. List mode em reconstruction of compton scatter camera images in 3-d. In *2000 IEEE Nuclear Science Symposium. Conference Record (Cat. No. 00CH37149)*, volume 2, pages 15–292. IEEE, 2000.

[6] Russell S Williford. Temporal gamma-ray spectrometry to quantify relative fissile material content. 2013.

[7] Lawrence Rabiner and Biinghwang Juang. An introduction to hidden markov models. *ieee assp magazine*, 3(1):4–16, 1986.

[8] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[9] Lekha Patel, Nils Gustafsson, Yu Lin, Raimund Ober, Ricardo Henriques, and Edward Cohen. A hidden markov model approach to characterizing the photo-switching behavior of fluorophores. *The annals of applied statistics*, 13(3):1397, 2019.

[10] Zoubin Ghahramani and Michael Jordan. Factorial hidden markov models. *Advances in neural information processing systems*, 8, 1995.

[11] A Tamhane and Dunlop Dunlop. *Statistics and Data Analysis: From Elementary to Intermediate*. Prentice-Hall, 2000.

[12] William Jay Conover. *Practical nonparametric statistics*, volume 350. john wiley & sons, 1999.

[13] Robert H Shumway and David S Stoffer. *Time series analysis and its applications (springer texts in statistics)*. Springer-Verlag, 2005.

[14] Dean J Mitchell, Lee Harding, Gregory G Thoreson, and Steven M Horne. Gadras detector response function. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2014.

[15] Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman. *Mining of massive data sets*. Cambridge university press, 2020.

[16] Vesselin Petkov and Maciej Zworski. Breit–wigner approximation and the distribution¶ of resonances. *Communications in mathematical physics*, 204:329–351, 1999.

[17] Nickos Papadatos. The characteristic function of the discrete cauchy distribution in memory of t. cacoullos. *Journal of Statistical Theory and Practice*, 16(3):47, 2022.

[18] Stanisław Węglarczyk. Kernel density estimation and its application. In *ITM web of conferences*, volume 23, page 00037. EDP Sciences, 2018.

[19] Francis Sahngun Nahm. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean journal of anesthesiology*, 75(1):25–36, 2022.

[20] Seong Ho Park, Jin Mo Goo, and Chan-Hee Jo. Receiver operating characteristic (roc) curve: practical review for radiologists. *Korean journal of radiology*, 5(1):11–18, 2004.

[21] David J Hand and Robert J Till. A simple generalisation of the area under the roc curve for multiple class classification problems. *Machine learning*, 45:171–186, 2001.

[22] Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez, and Markus Müller. proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC bioinformatics*, 12:1–8, 2011.

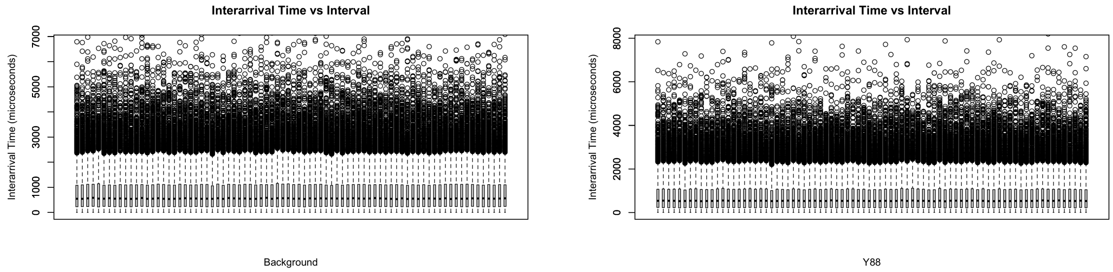
[23] Katarzyna Stąpor. Evaluating and comparing classifiers: Review, some recommendations and limitations. In *Proceedings of the 10th International Conference on Computer Recognition Systems CORES 2017 10*, pages 12–21. Springer, 2018.

[24] Thomas G Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, 10(7):1895–1923, 1998.

[25] Eve Richardson, Raphael Trevizani, Jason A Greenbaum, Hannah Carter, Morten Nielsen, and Bjoern Peters. The receiver operating characteristic curve accurately assesses imbalanced datasets. *Patterns*, 2024.

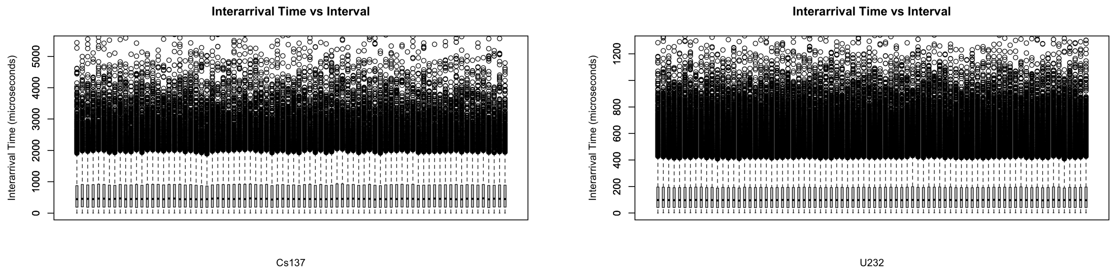
APPENDIX A. Full Results from Data Analysis and Classifier Experiments

A.1. Box Plots



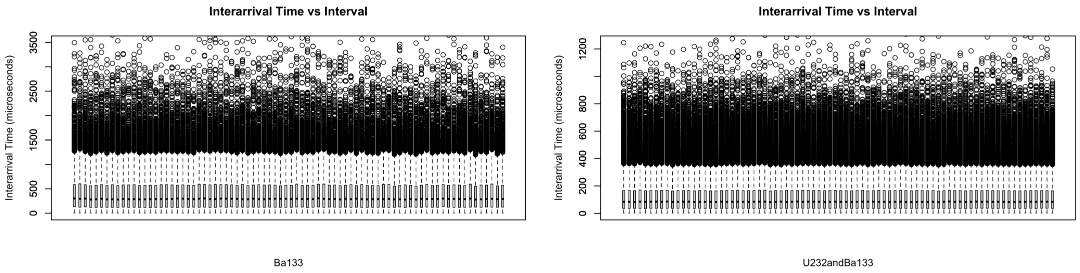
(a) Distributions of Interarrival Times of Background on July 2022

(b) Distributions of Interarrival Times of Y88 on July 2022



(c) Distributions of Interarrival Times of Cs137 on July 2022

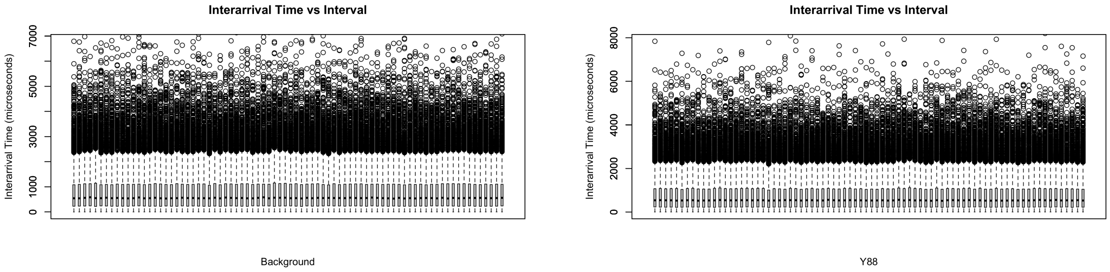
(d) Distributions of Interarrival Times of U232 on July 2022



(e) Distributions of Interarrival Times of Ba133 on July 2022

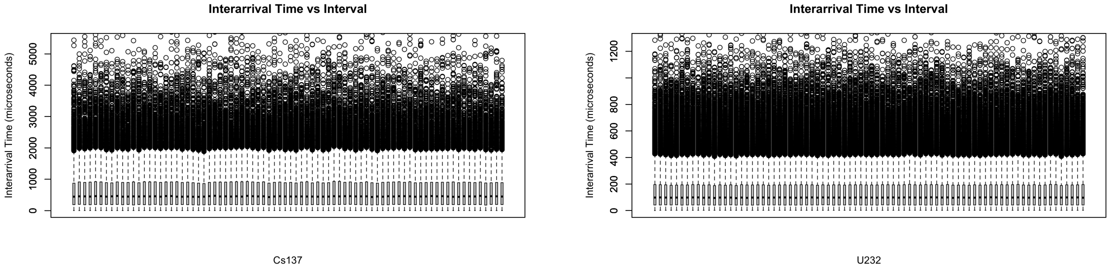
(f) Distributions of Interarrival Times of U232 & Ba133 on July 2022

Figure A-1. Figures A-1a, A-1b, A-1c, A-1d, A-1e, and A-1f display the change in the distribution of interarrival times associated with background, Y88, Cs137, U232, Ba133, and the combination of U232 and Ba133 as the 20s window over which the sample was collected changed. The list-mode data associated with these plots was collected in July 2022.



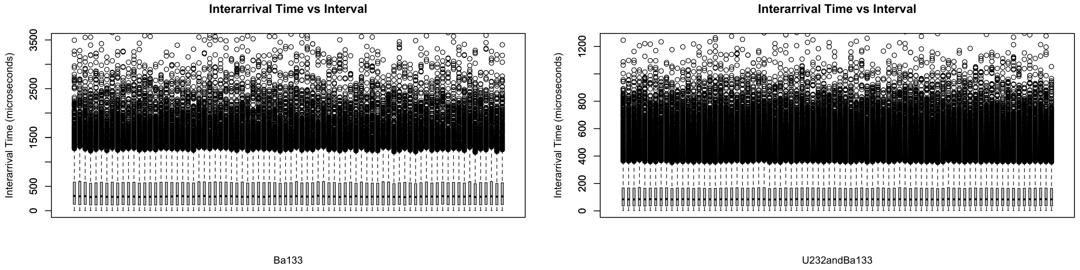
(a) Distributions of Interarrival Times of Background on July 2022

(b) Distributions of Interarrival Times of Y88 on July 2022



(c) Distributions of Interarrival Times of Cs137 on July 2022

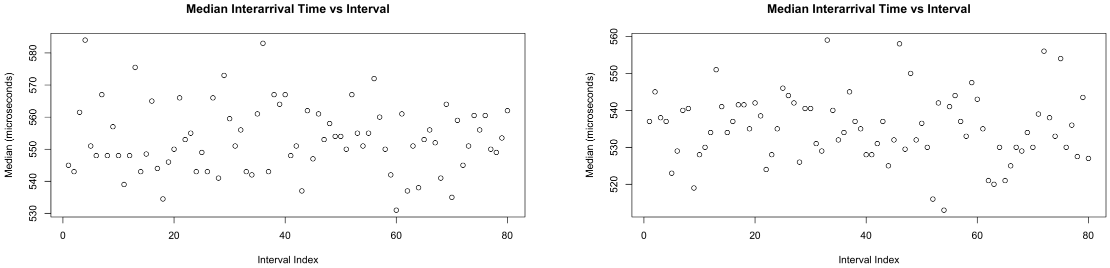
(d) Distributions of Interarrival Times of U232 on July 2022



(e) Distributions of Interarrival Times of Ba133 on July 2022

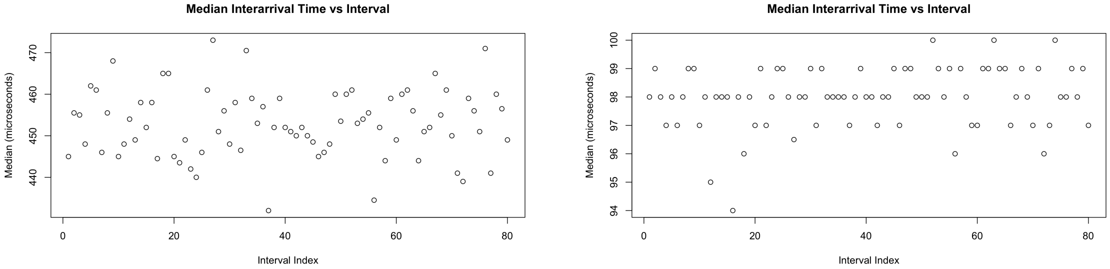
(f) Distributions of Interarrival Times of U232 & Ba133 on July 2022

Figure A-2. Figures A-2a, A-2b, A-2c, A-2d, A-2e, and A-2f display the change in the distribution of interarrival times associated with background, Y88, Cs137, U232, Ba133, and the combination of U232 and Ba133 as the size of the sampling window changed from 20s (left) to 0.2s (right). The list-mode data associated with these plots was collected in July 2022.



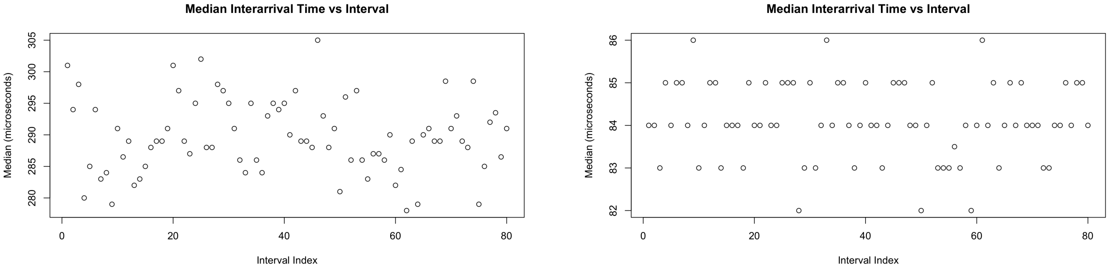
(a) Median of Interarrival Times of Background on July 2022

(b) Median of Interarrival Times of Y88 on July 2022



(c) Median of Interarrival Times of Cs137 on July 2022

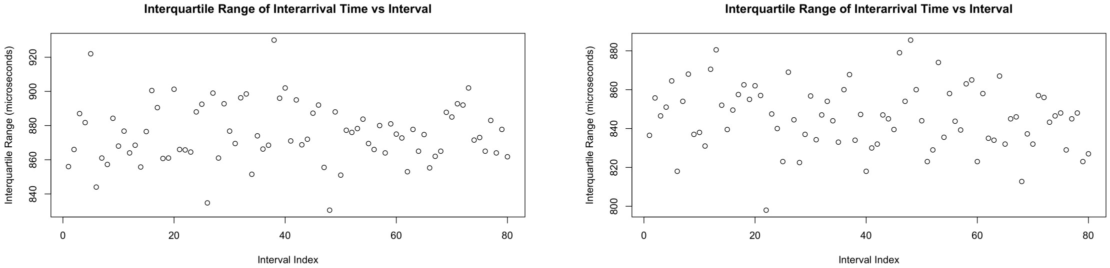
(d) Median of Interarrival Times of U232 on July 2022



(e) Median of Interarrival Times of Ba133 on July 2022

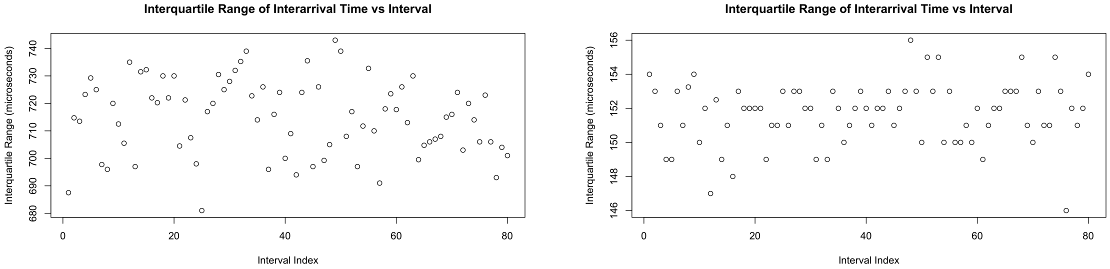
(f) Median of Interarrival Times of U232 & Ba133 on July 2022

Figure A-3. Figures A-3a, A-3b, A-3c, A-3d, A-3e, and A-3f display the change in the median of the distribution of interarrival times associated with background, Y88, Cs137, U232, Ba133, and the combination of U232 and Ba133 as the sampling window changes. The list-mode data associated with these plots was collected in July 2022.



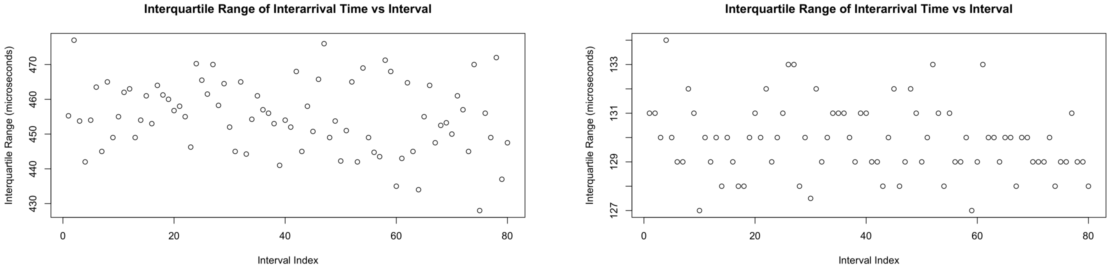
(a) IQR of Interarrival Times of Background on July 2022

(b) IQR of Interarrival Times of Y88 on July 2022



(c) IQR of Interarrival Times of Cs137 on July 2022

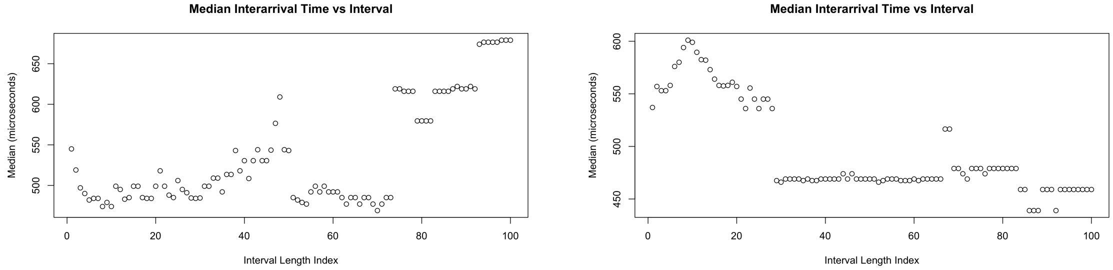
(d) IQR of Interarrival Times of U232 on July 2022



(e) IQR of Interarrival Times of Ba133 on July 2022

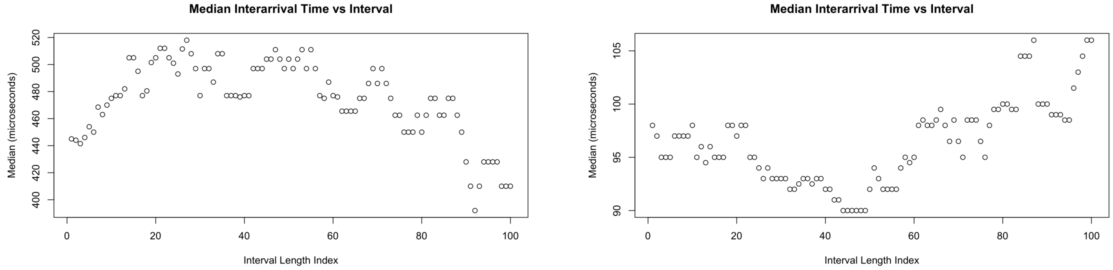
(f) IQR of Interarrival Times of U232 & Ba133 on July 2022

Figure A-4. Figures A-4a, A-4b, A-4c, A-4d, A-4e, and A-4f display the change in the interquartile range of the distribution of interarrival times associated with background, Y88, Cs137, U232, Ba133, and the combination of U232 and Ba133 as the sampling window changes. The list-mode data associated with these plots was collected in July 2022.



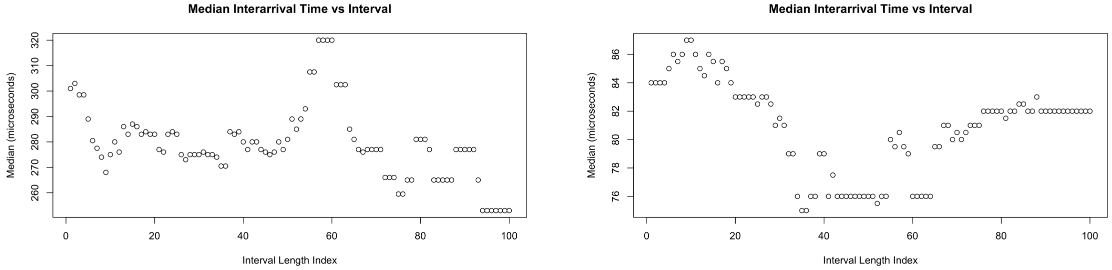
(a) Median of Interarrival Times of Background on July 2022

(b) Median of Interarrival Times of Y88 on July 2022



(c) Median of Interarrival Times of Cs137 on July 2022

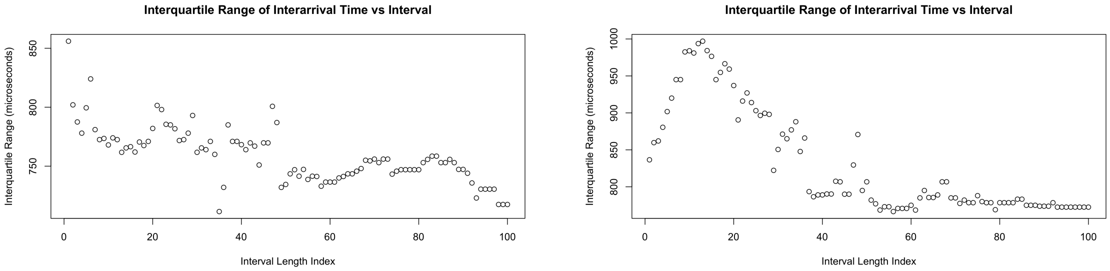
(d) Median of Interarrival Times of U232 on July 2022



(e) Median of Interarrival Times of Ba133 on July 2022

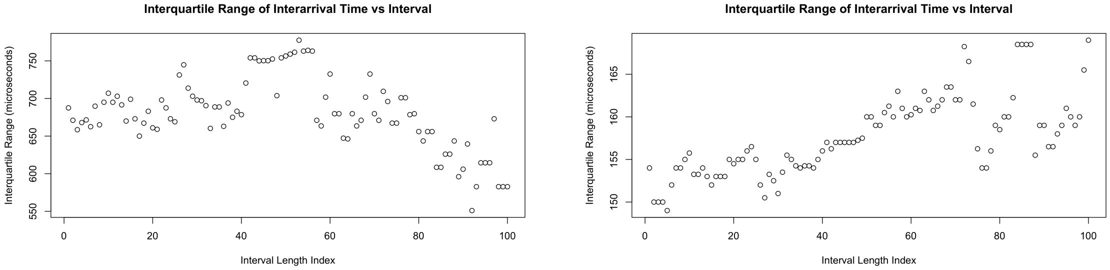
(f) Median of Interarrival Times of U232 & Ba133 on July 2022

Figure A-5. Figures A-5a, A-5b, A-5c, A-5d, A-5e, and A-5f display the change in the median of the distribution of interarrival times associated with background, Y88, Cs137, U232, Ba133, and the combination of U232 and Ba133 as the length of the sampling window changes from 20s (left) to 0.2s (right). The list-mode data associated with these plots was collected in July 2022.



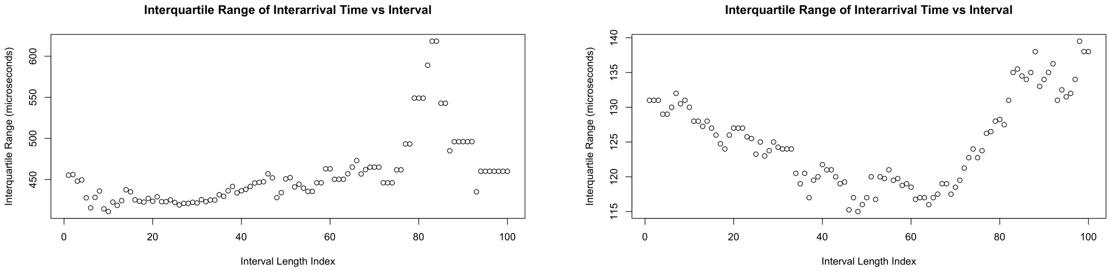
(a) IQR of Interarrival Times of Background on July 2022

(b) IQR of Interarrival Times of Y88 on July 2022



(c) IQR of Interarrival Times of Cs137 on July 2022

(d) IQR of Interarrival Times of U232 on July 2022

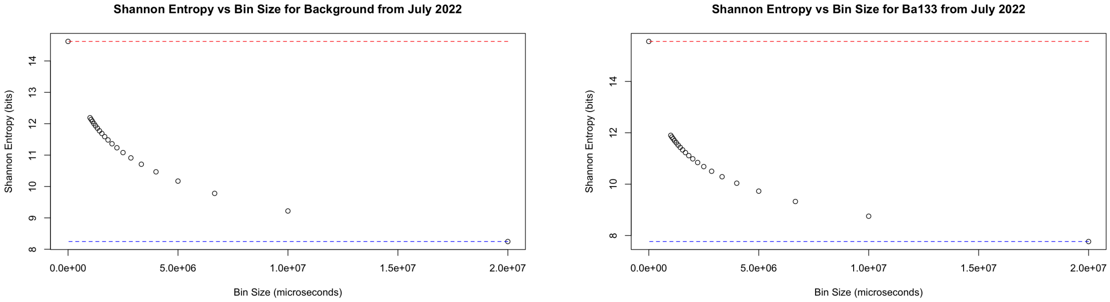


(e) IQR of Interarrival Times of Ba133 on July 2022

(f) IQR of Interarrival Times of U232 & Ba133 on July 2022

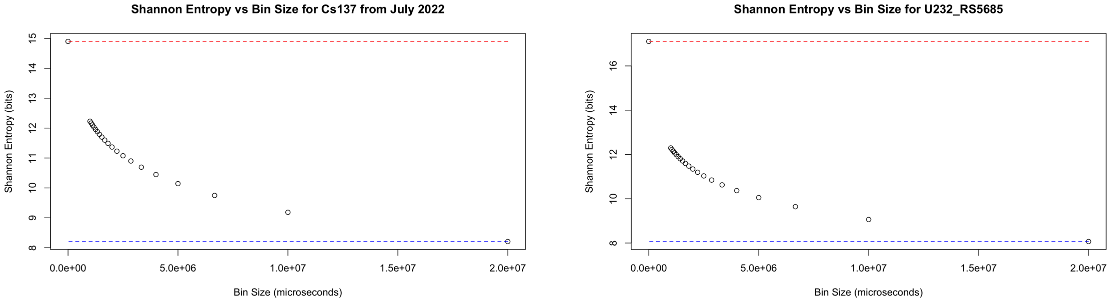
Figure A-6. Figures A-6a, A-6b, A-6c, A-6d, A-6e, and A-6f display the change in the interquartile range of the distribution of interarrival times associated with background, Y88, Cs137, U232, Ba133, and the combination of U232 and Ba133 as the length of the sampling window changes from 20s (left) to 0.2s (right). The list-mode data associated with these plots was collected in July 2022.

A.2. Shannon Entropy Plots



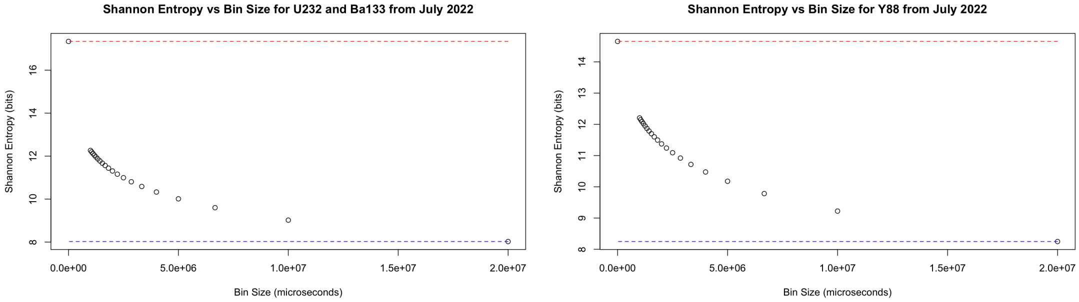
(a) Shannon Entropy for Binning of Background on July 2022

(b) Shannon Entropy for Binning of Ba133 on July 2022



(c) Shannon Entropy for Binning of Cs137 on July 2022

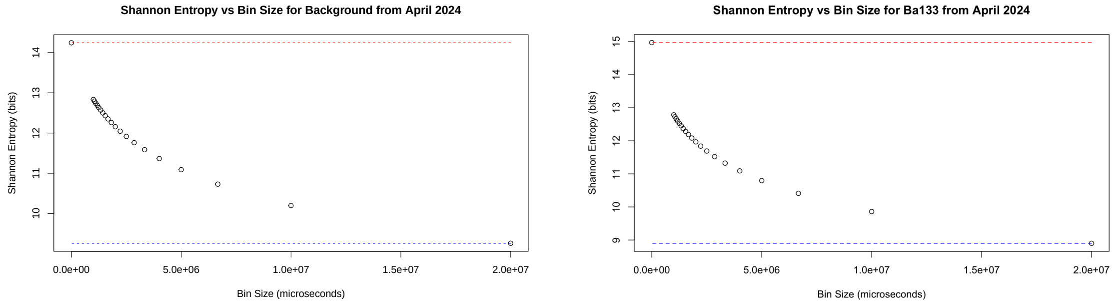
(d) Shannon Entropy for Binning of U232 on July 2022



(e) Shannon Entropy for Binning of U232 & Ba133 on July 2022

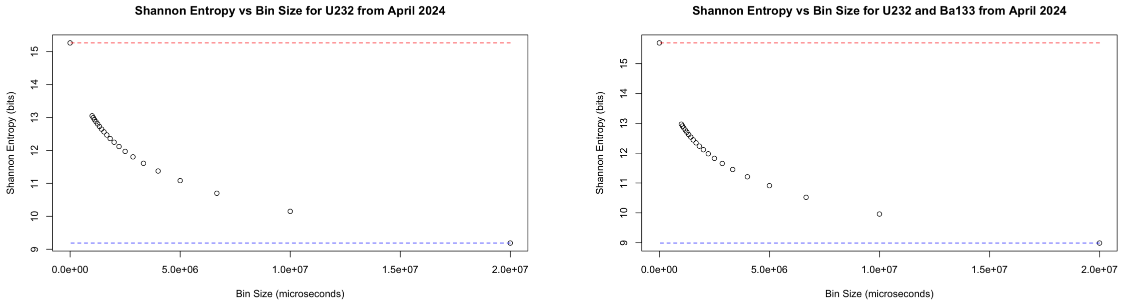
(f) Shannon Entropy for Binning of Y88 on July 2022

Figure A-7. Figures A-7a, A-7b, A-7c, A-7d, A-7e, and A-7f display the sShannon entropy associated with the binning of background, Ba133, Cs137, U232, the combination of U232 and Ba133, and Y88, respectively. The list-mode data associated with these plots was collected in July 2022.



(a) Shannon Entropy for Binning of Background on April 2024

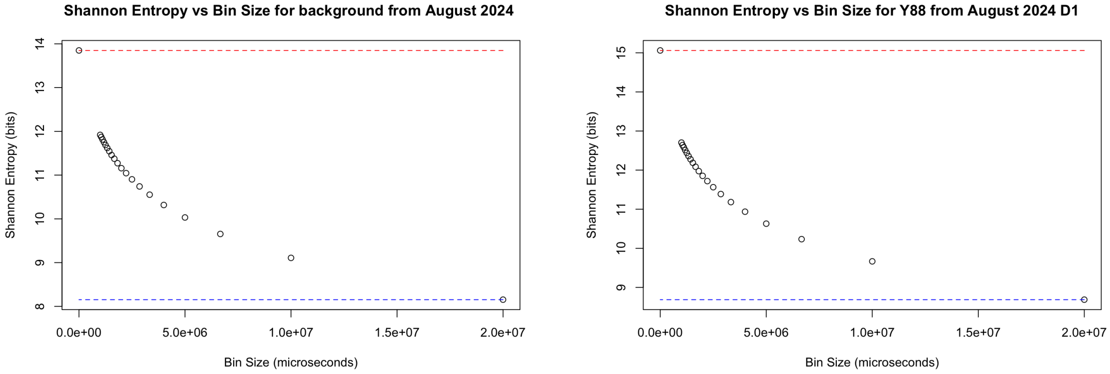
(b) Shannon Entropy for Binning of Ba133 on April 2024



(c) Shannon Entropy for Binning of U232 on April 2024

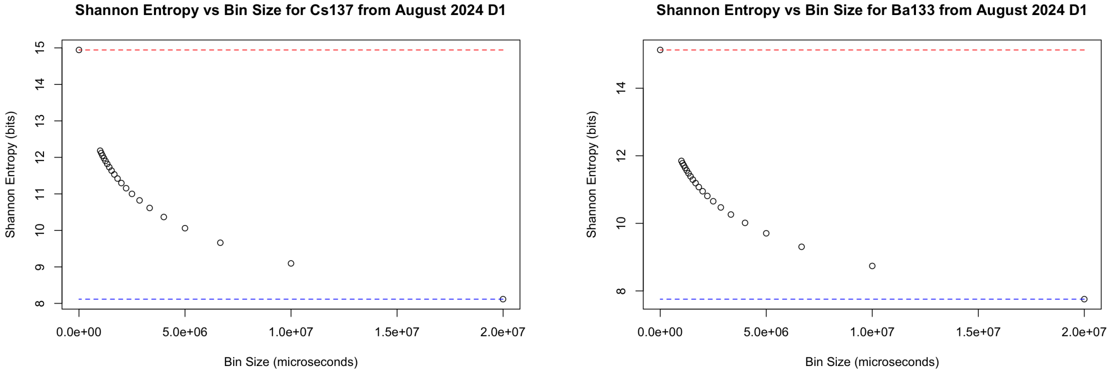
(d) Shannon Entropy for Binning of U232 & Ba133 on April 2024

Figure A-8. Figures A-8a, A-8b, A-8c, and A-8d display the Shannon entropy associated with the binning of background, Ba133, U232, and the combination of U232 and Ba133, respectively. The list-mode data associated with these plots was collected in April 2024.



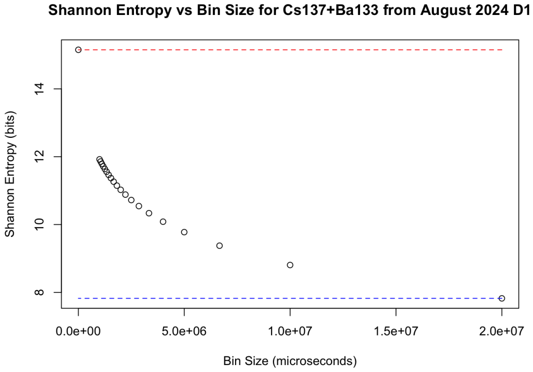
(a) Shannon Entropy for Binning of Background on August 2024

(b) Shannon Entropy for Binning of Y88 on August 2024 D1



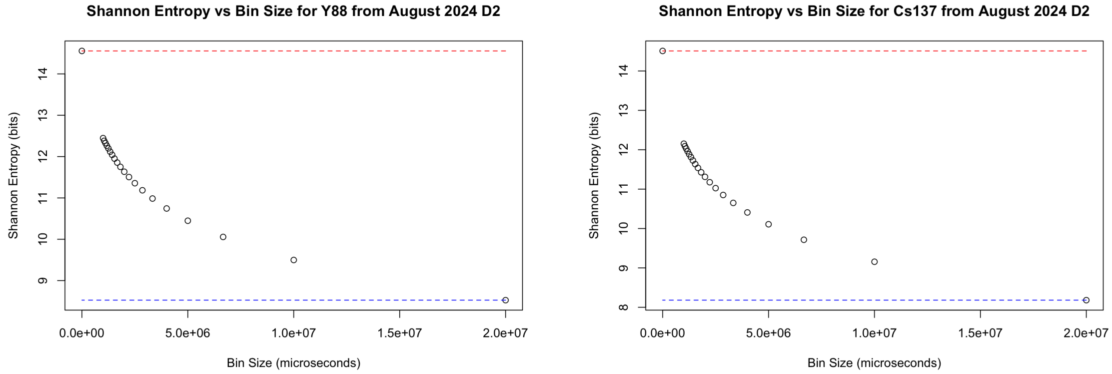
(c) Shannon Entropy for Binning of Cs137 on August 2024 D1

(d) Shannon Entropy for Binning of Ba133 on August 2024 D1



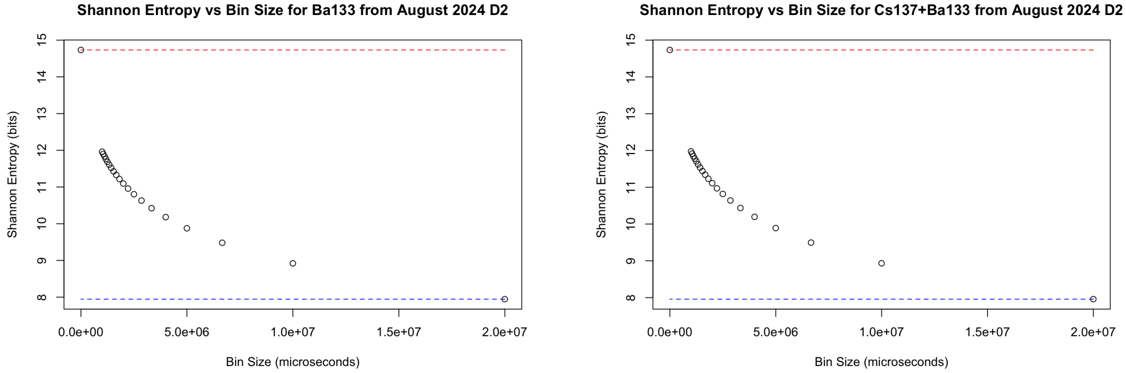
(e) Shannon Entropy for Binning of Cs137 & Ba133 on August 2024 D1

Figure A-9. Figures A-9a, A-9b, A-9c, A-9d, A-9e display the Shannon entropy associated with the binning of background, Y88, Cs137, Ba133, and the combination of Cs137 and Ba133, respectively. The list-mode data associated with these plots was collected in August 2024 with D1 representing the set of distances at which the data was collected.



(a) Shannon Entropy for Binning of Y88 on August 2024 D2

(b) Shannon Entropy for Binning of Cs137 on August 2024 D2

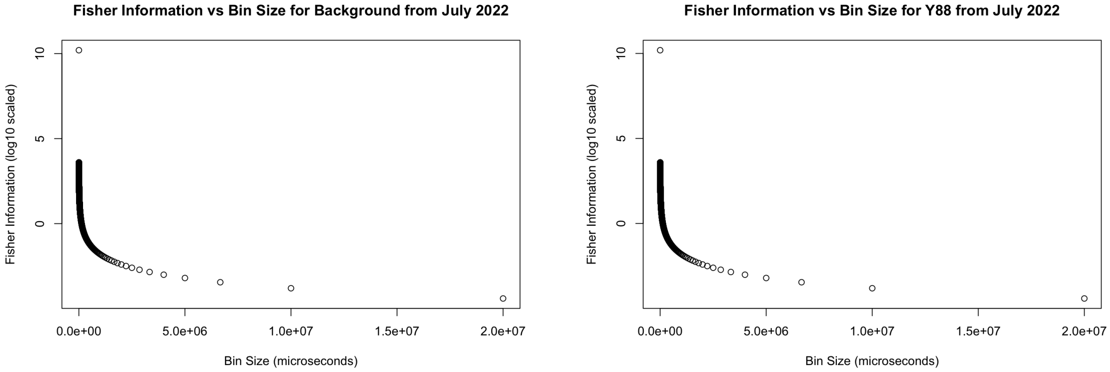


(c) Shannon Entropy for Binning of Ba133 on August 2024 D2

(d) Shannon Entropy for Binning of Cs137 & Ba133 on August 2024 D2

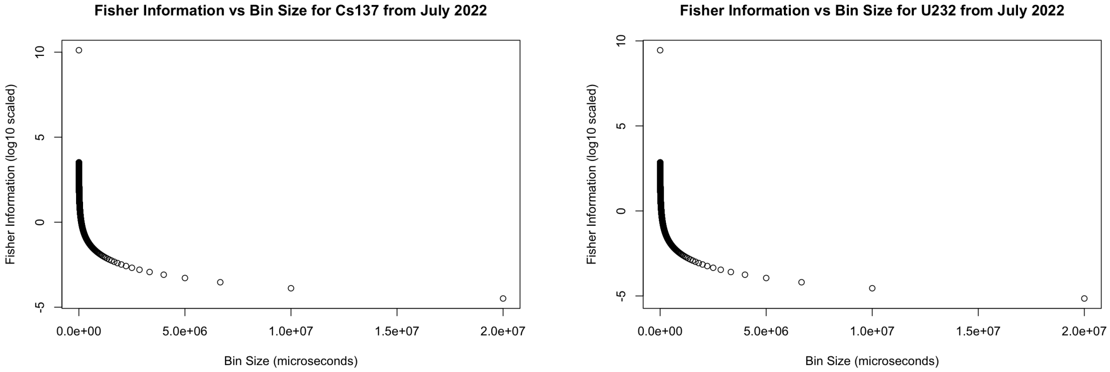
Figure A-10. Figures A-10a, A-10b, A-10c, and A-10d display the Shannon entropy associated with the binning of Y88, Cs137, Ba133, and the combination of Cs137 and Ba133, respectively. The list-mode data associated with these plots was collected in August 2024 with D2 representing the set of distances at which the data was collected.

A.3. Fisher Information Plots



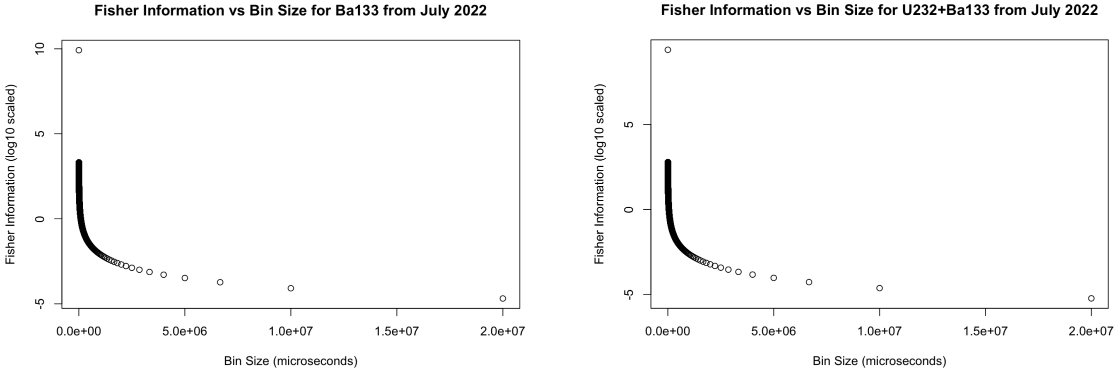
(a) Fisher information for the rate parameter of Background on July 2022

(b) Fisher information for the rate parameter of Y88 on July 2022



(c) Fisher information for the rate parameter of Cs137 on July 2022

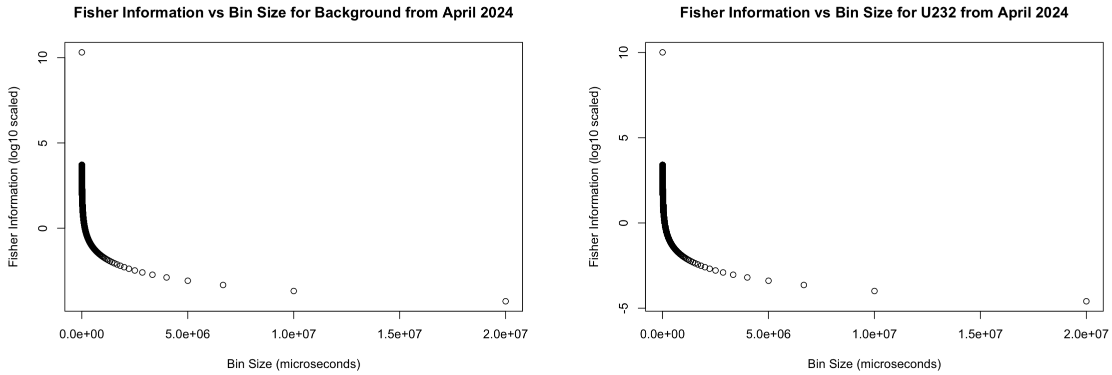
(d) Fisher information for the rate parameter of U232 on July 2022



(e) Fisher information for the rate parameter of Ba133 on July 2022

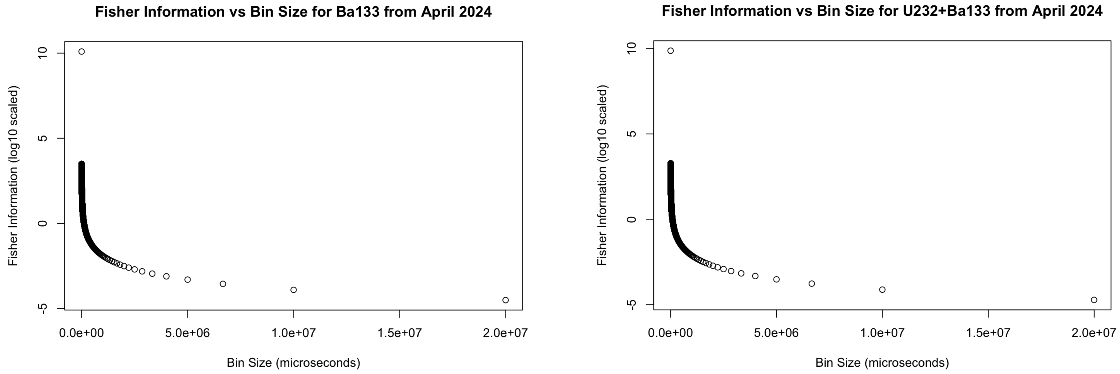
(f) Fisher information for the rate parameter of U232&Ba133 on July 2022

Figure A-11. Figures A-11a, A-11b, A-11c, A-11d, A-11e, and A-11f display the change in the Fisher information associated with the rate parameter as the size of each bin is increased. The list-mode data associated with these plots was collected in July 2022.



(a) Fisher information for the rate parameter of Background on April 2024

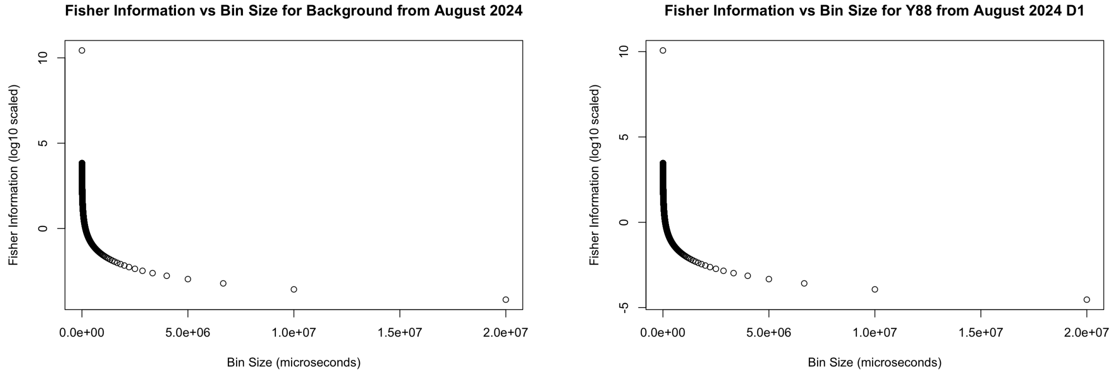
(b) Fisher information for the rate parameter of U232 on April 2024



(c) Fisher information for the rate parameter of Ba133 on April 2024

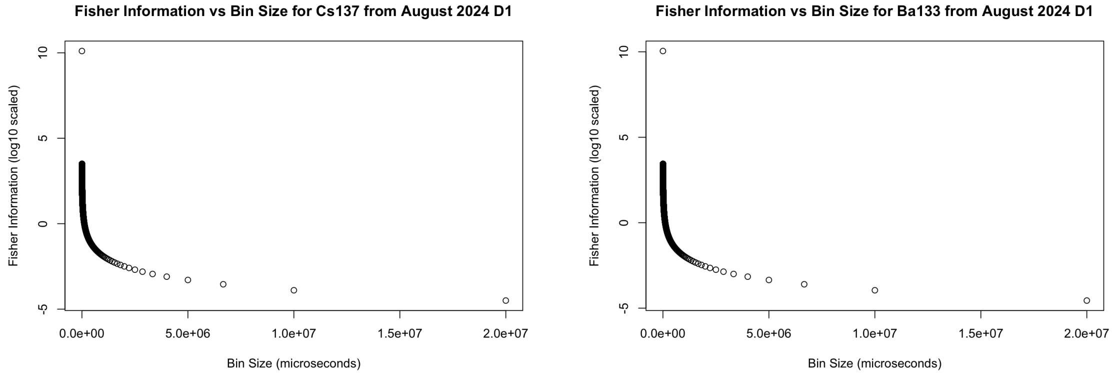
(d) Fisher information for the rate parameter of U232&Ba133 on April 2024

Figure A-12. Figures A-12a, A-12b, A-12c, and A-12d display the change in the Fisher information associated with the rate parameter as the size of each bin is increased. The list-mode data associated with these plots was collected in April 2024.



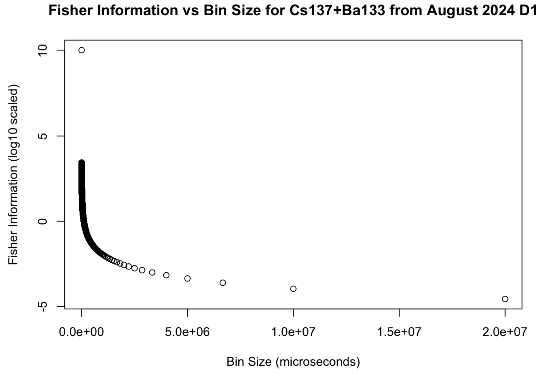
(a) Fisher information for the rate parameter of Background on August 2024

(b) Fisher information for the rate parameter of Y88 on August 2024 D1



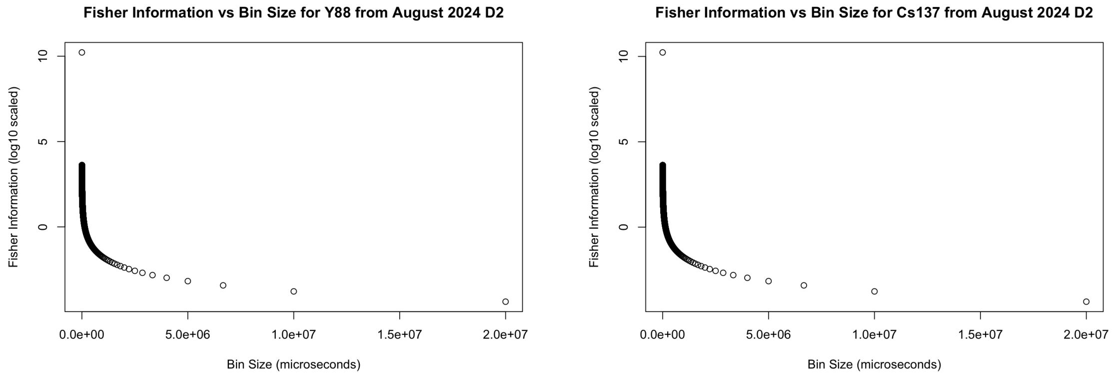
(c) Fisher information for the rate parameter of Cs137 on August 2024 D1

(d) Fisher information for the rate parameter of Ba133 on August 2024 D1



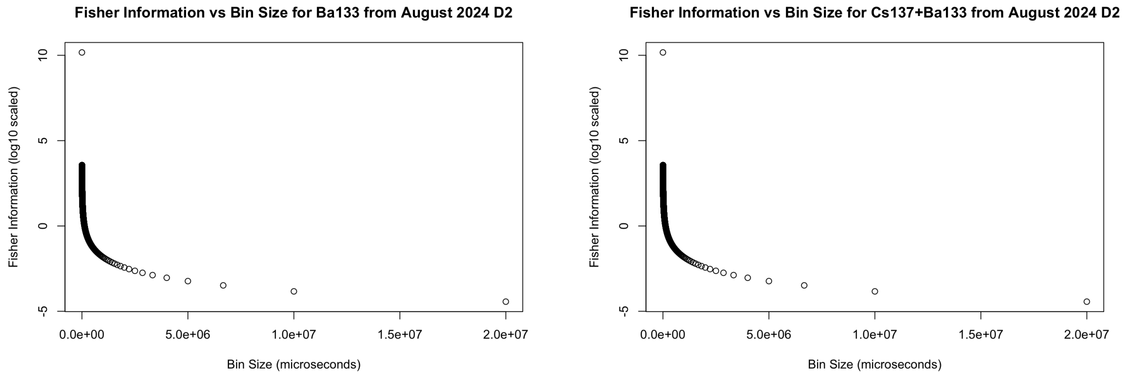
(e) Fisher information for the rate parameter of Cs137&Ba133 on August 2024 D1

Figure A-13. Figures A-13a, A-13b, A-13c, A-13d, and A-13e display the change in the Fisher information associated with the rate parameter as the size of each bin is increased. The list-mode data associated with these plots was collected in August 2024. The source data was collected at the distances associated with dataset D1.



(a) Fisher information for the rate parameter of Y88 on August 2024 D2

(b) Fisher information for the rate parameter of Cs137 on August 2024 D2

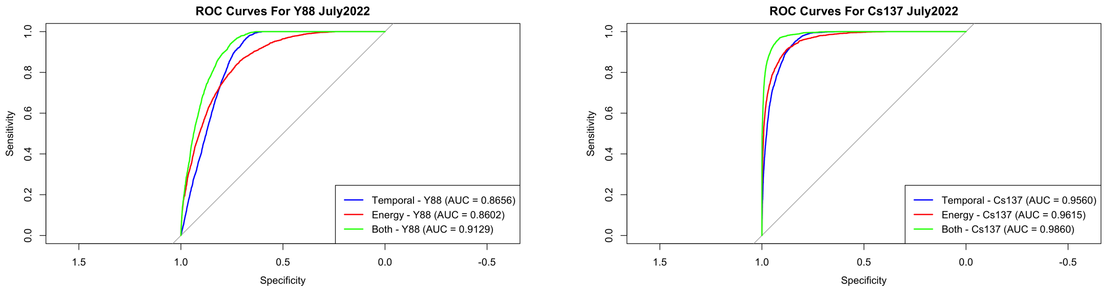


(c) Fisher information for the rate parameter of Ba133 on August 2024 D2

(d) Fisher information for the rate parameter of Cs137&Ba133 on August 2024 D2

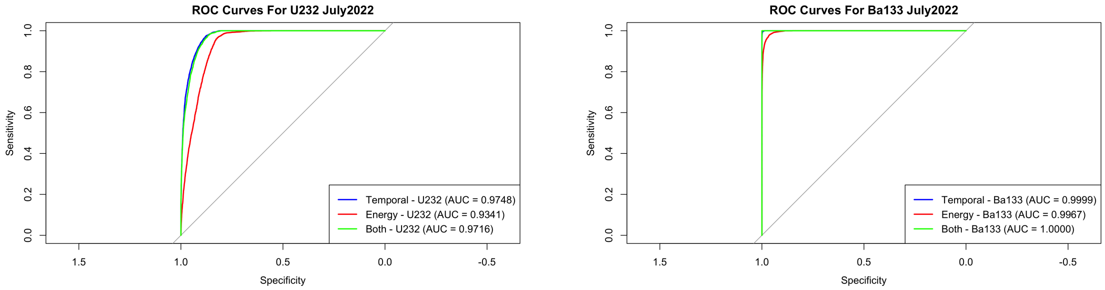
Figure A-14. Figures A-13a, A-13b, A-13c, and A-13d display the change in the Fisher information associated with the rate parameter as the size of each bin is increased. The list-mode data associated with these plots was collected in August 2024. The source data was collected at the distances associated with dataset D2.

A.4. ROC curves and AUC values



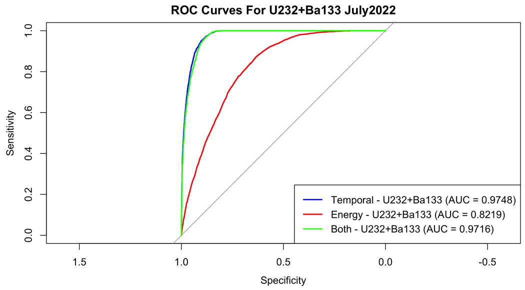
(a) Experiment 1: ROC for model performance on Y88

(b) Experiment 1: ROC for model performance on Cs137



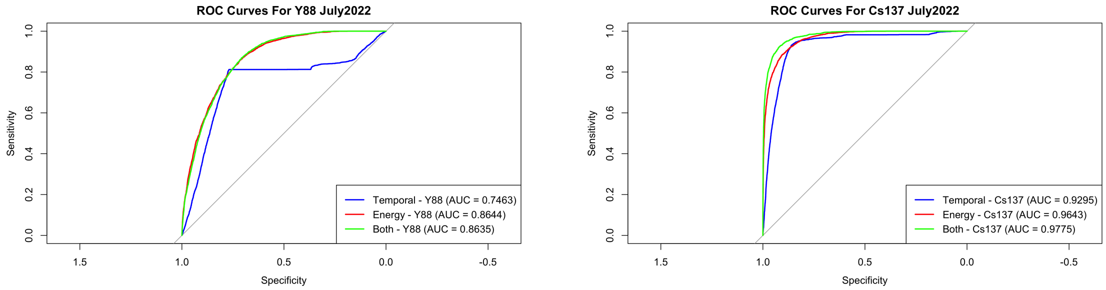
(c) Experiment 1: ROC for model performance on U232

(d) Experiment 1: ROC for model performance on Ba133



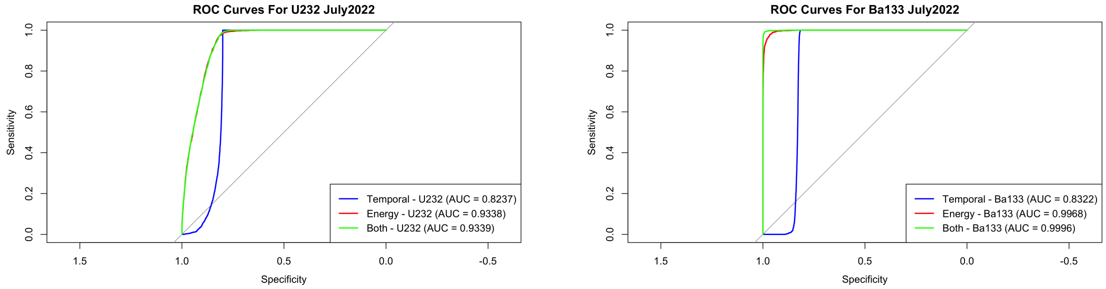
(e) Experiment 1: ROC for model performance on U232 & Ba133

Figure A-15. Figures A-15a, A-15b, A-15c, A-15d, and A-15e display the ROCs associated with the performance of the time, energy, and energy & time models on the different sources. The list-mode data associated with these plots was collected in July 2022.



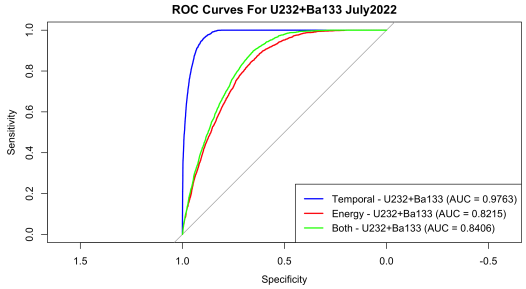
(a) Experiment 1: ROC for α-model performance on Y88

(b) Experiment 1: ROC for α-model performance on Cs137



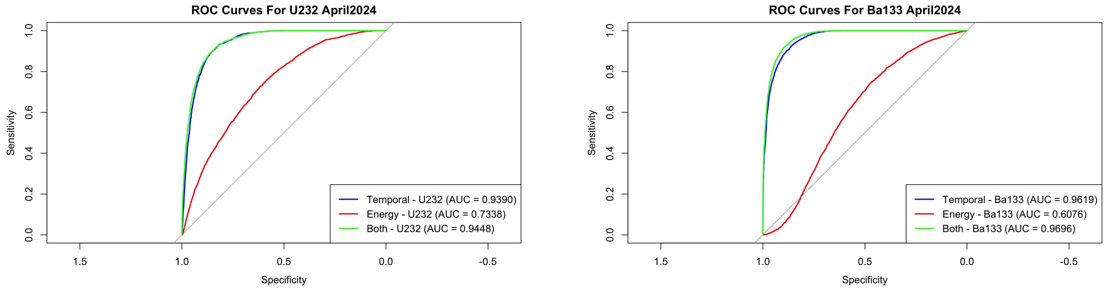
(c) Experiment 1: ROC for α-model performance on U232

(d) Experiment 1: ROC for α-model performance on Ba133



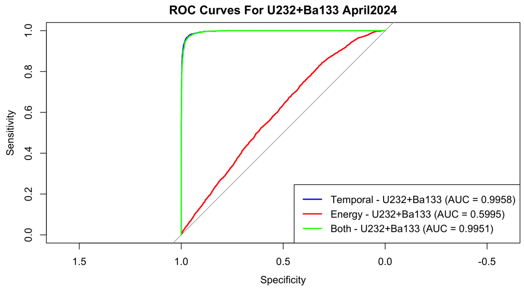
(e) Experiment 1: ROC for α-model performance on U232 & Ba133

Figure A-16. Figures A-16a, A-16b, A-16c, A-16d, and A-16e display the ROCs associated with the performance of the time, energy, and energy & time models on the different sources using the extended α model. The list-mode data associated with these plots was collected in July 2022.



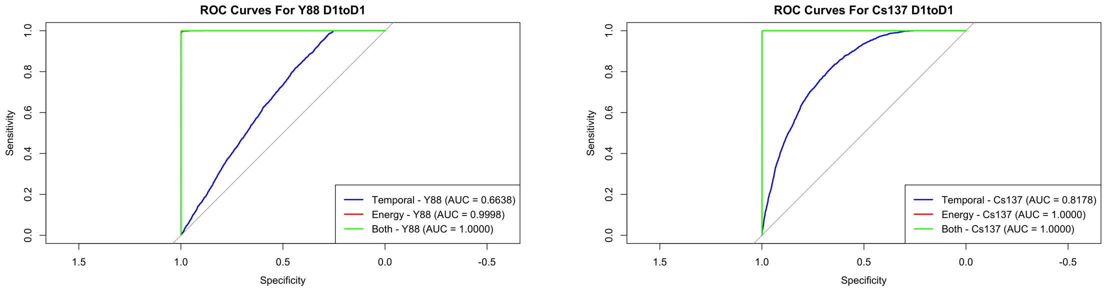
(a) Experiment 2: ROC for model performance on U232

(b) Experiment 2: ROC for model performance on Ba133 on April 2024



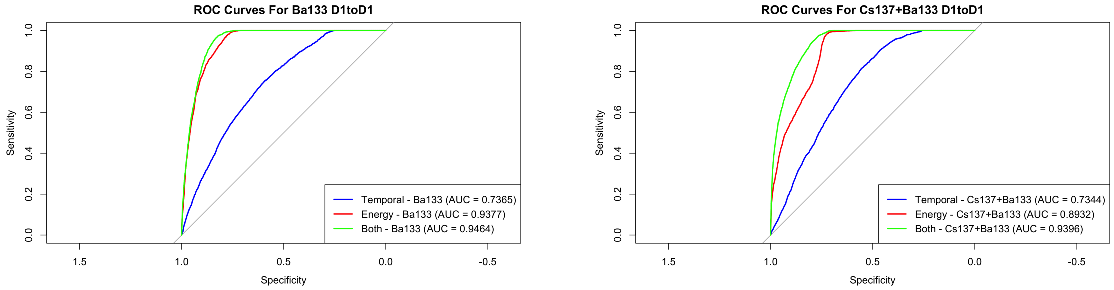
(c) Experiment 2: ROC for model performance on U232 & Ba133

Figure A-17. Figures A-17a, A-17b, and A-17c display the ROCs associated with the performance of the time, energy, and energy & time models on the different sources. The list-mode data associated with these plots was collected in April 2024.



(a) Experiment 3: ROC for model performance on Y88

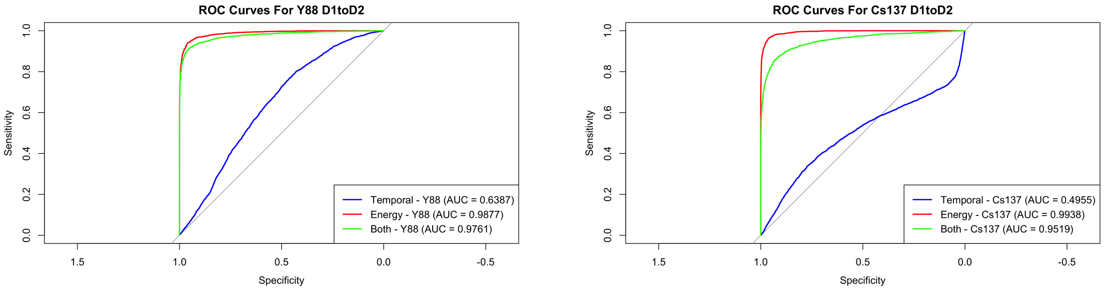
(b) Experiment 3: ROC for model performance on Cs137



(c) Experiment 3: ROC for model performance on Ba133

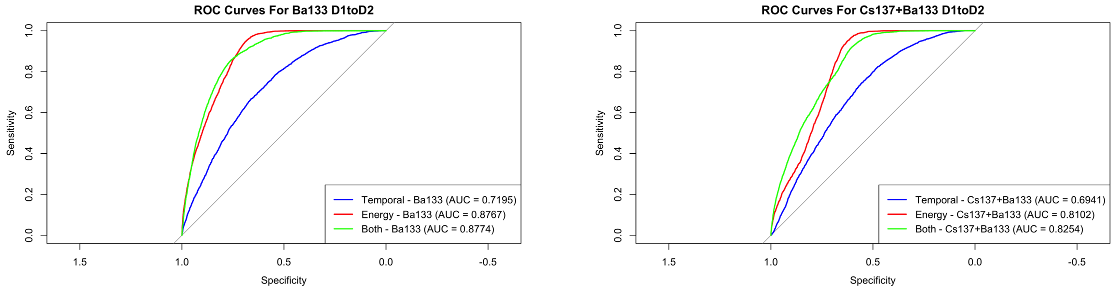
(d) Experiment 3: ROC for model performance on Cs137 & Ba133

Figure A-18. Figures A-18a, A-18b, A-18c, and A-18d display the ROCs associated with the performance of the time, energy, and energy & time models on the different sources. The test and training data associated with these plots is list-mode data that was collected in August 2024 at the D1 distances.



(a) Experiment 4: ROC for model performance on Y88

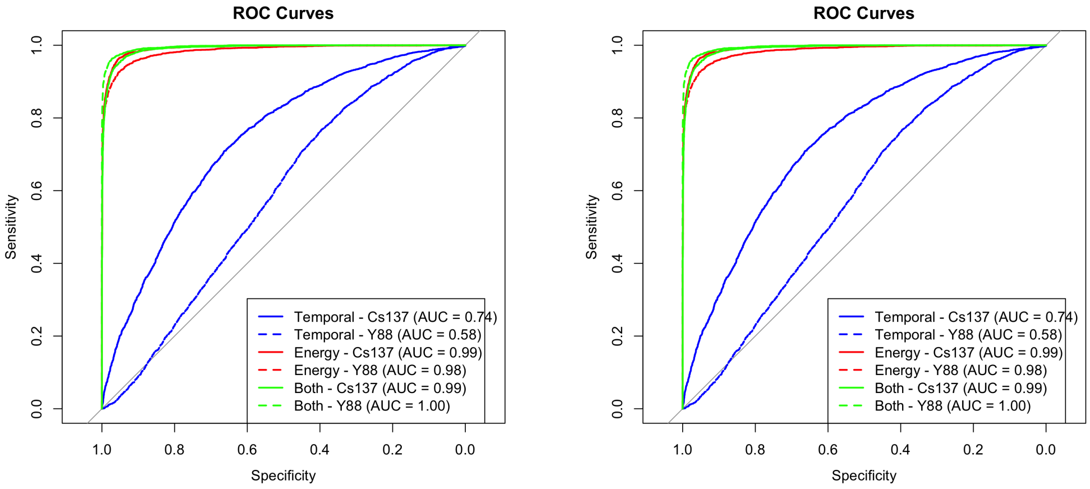
(b) Experiment 4: ROC for model performance on Cs137



(c) Experiment 4: ROC for model performance on Ba133

(d) Experiment 4: ROC for model performance on Cs137 & Ba133

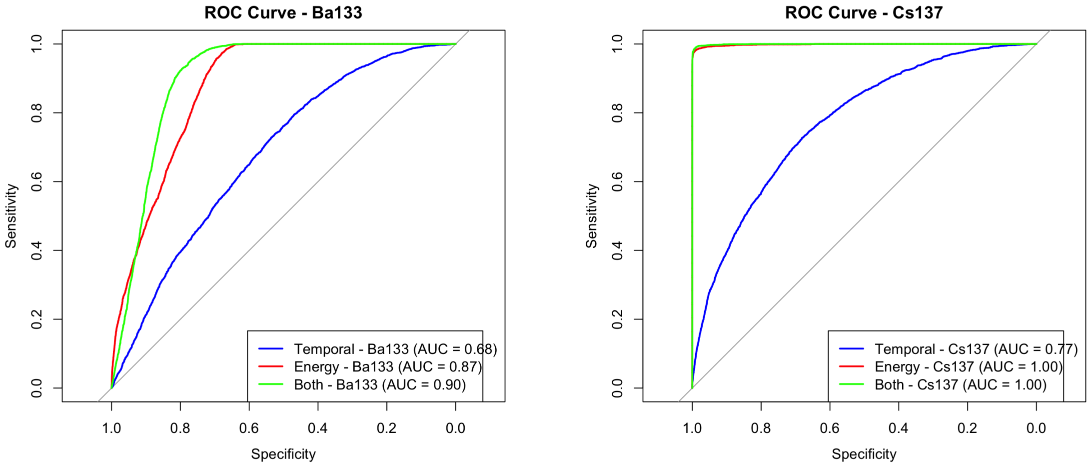
Figure A-19. Figures A-19a, A-19b, A-19c, and A-19d display the ROCs associated with the performance of the time, energy, and energy & time models on the different sources. The test data associated with these plots is list-mode data that was collected in August 2024 at the D2 distances whereas the training set was collected at the D1 distances.



(a) Experiment 4: D2 to D1 ROC for α-model performance on Y88, Cs137 and Ba133

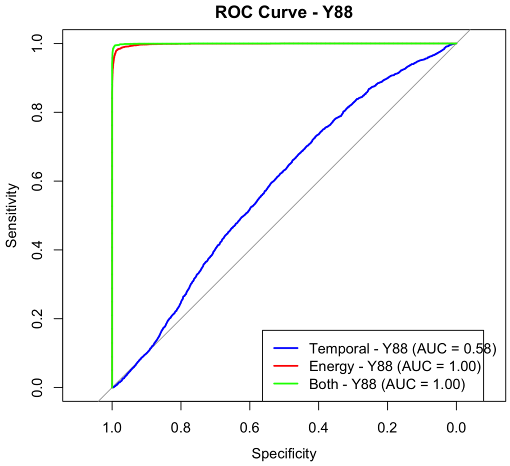
(b) Experiment 4: D1 to D2 ROC for α-model performance on on Y88, Cs137 and Ba133

Figure A-20. Figures A-20a and A-20b display the ROCs associated with the performance of the time, energy, and energy & time models on the different sources with the inclusion of α. The test data associated with these plots is list-mode data that was collected in August 2024 at the D1/D2 distances whereas the training set was collected at the D2/D1 distances.



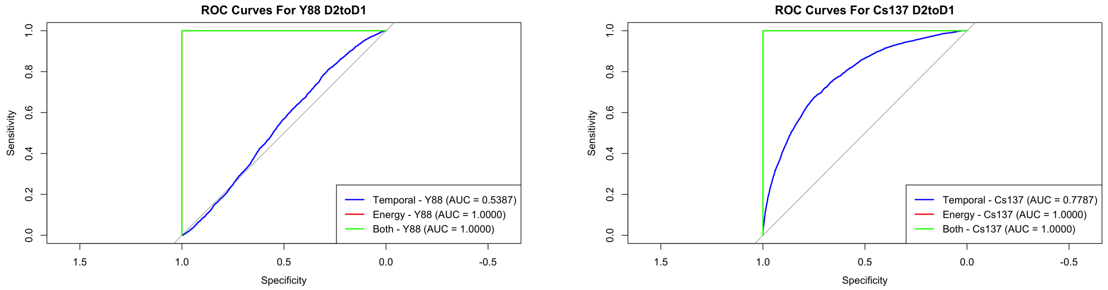
(a) Experiment 4: Mixed ROC for α-model performance on Ba133

(b) Experiment 4: Mixed ROC for α-model performance on Cs137 & Ba133



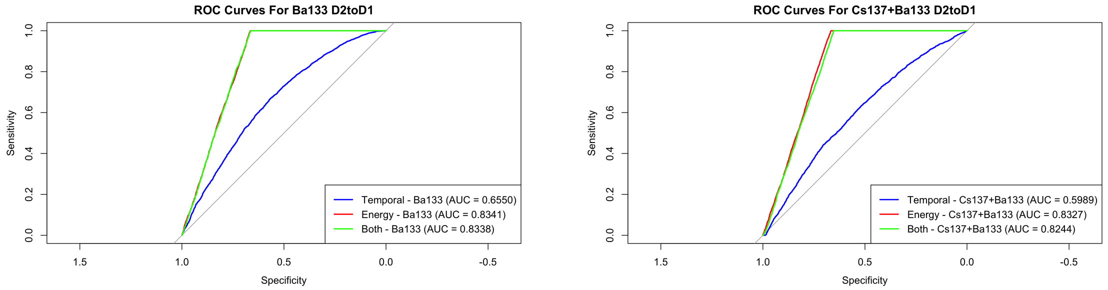
(c) Experiment 4: Mixed ROC for α-model performance on Y88

Figure A-21. Figures A-21a, A-21b and A-21c display the ROCs associated with the performance of the time, energy, and energy & time models on the different sources with the inclusion of α. The test and training data associated with these plots is list-mode data that was collected in August 2024 using the mixed distances.



(a) Experiment 5: ROC for model performance on Y88

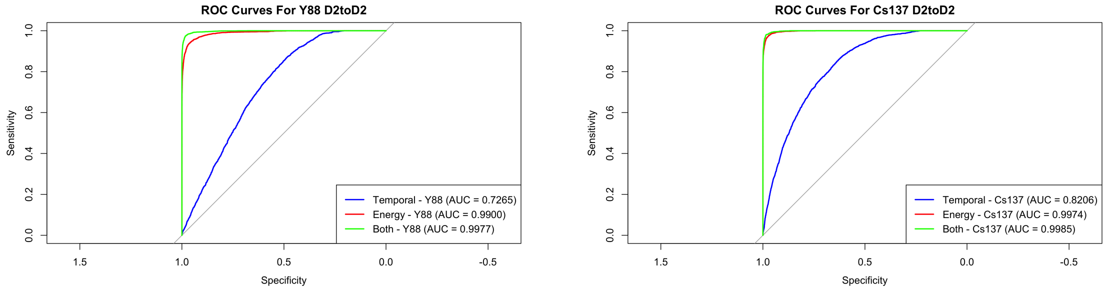
(b) Experiment 5: ROC for model performance on Cs137



(c) Experiment 5: ROC for model performance on Ba133

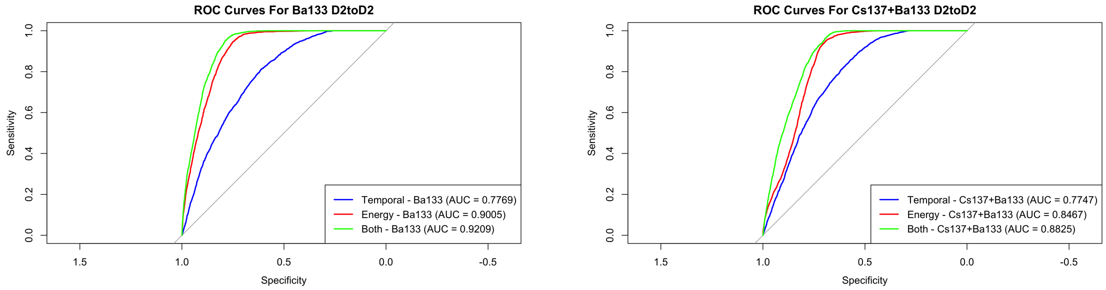
(d) Experiment 5: ROC for model performance on Cs137 & Ba133

Figure A-22. Figures A-22a, A-22b, A-22c, and A-22d display the ROCs associated with the performance of the time, energy, and energy & time models on the different sources. The test data associated with these plots is list-mode data that was collected in August 2024 at the D1 distances whereas the training set was collected at the D2 distances.



(a) Experiment 6: ROC for model performance on Y88

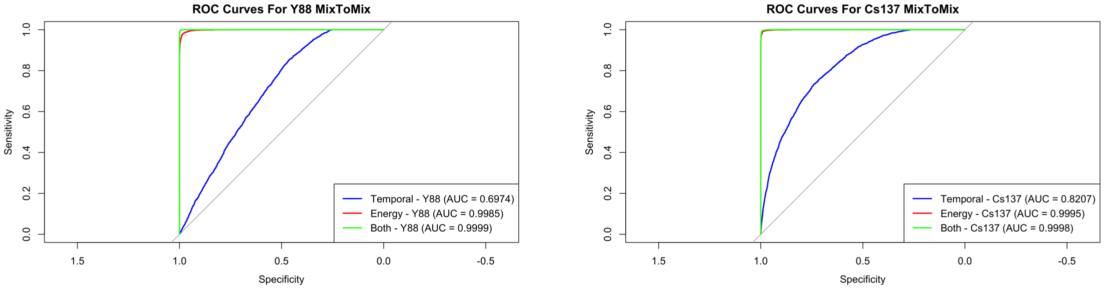
(b) Experiment 6: ROC for model performance on Cs137



(c) Experiment 6: ROC for model performance on Ba133

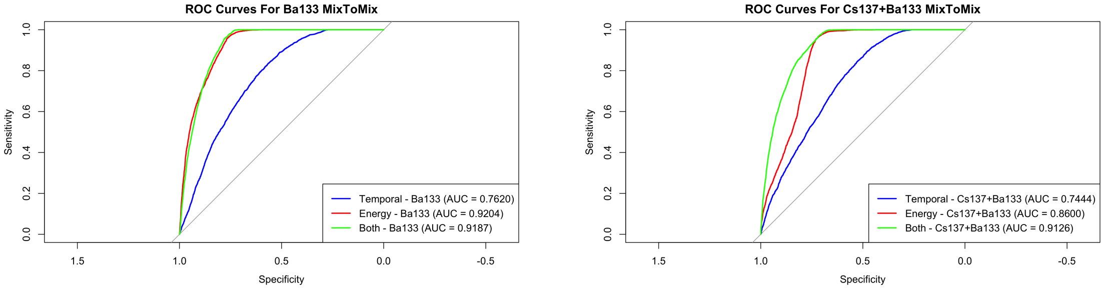
(d) Experiment 6: ROC for model performance on Cs137 & Ba133

Figure A-23. Figures A-23a, A-23b, A-23c, and A-23d display the ROCs associated with the performance of the time, energy, and energy & time models on the different sources. The test and training data associated with these plots is list-mode data that was collected in August 2024 at the D2 distances.



(a) Experiment 7: ROC for model performance on Y88

(b) Experiment 7: ROC for model performance on Cs137

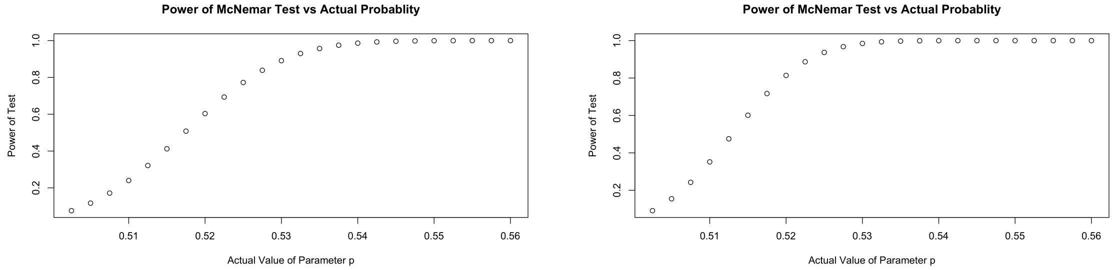


(c) Experiment 7: ROC for model performance on Ba133

(d) Experiment 7: ROC for model performance on Cs137 & Ba133

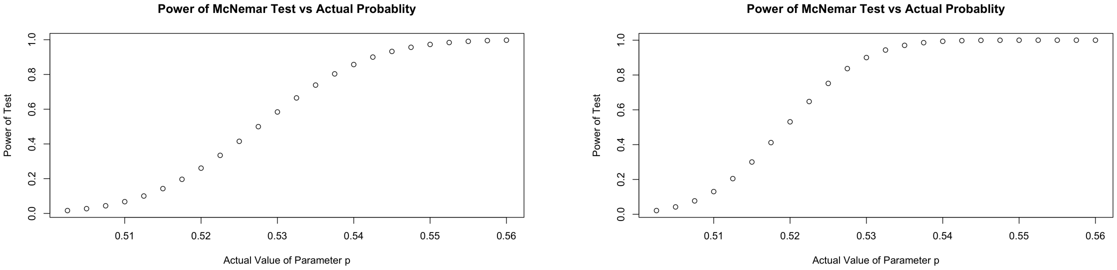
Figure A-24. Figures A-24a, A-24b, A-24c, and A-24d display the ROCs associated with the performance of the time, energy, and energy & time models on the different sources. The test and training data associated with these plots is a mixture of list-mode data that was collected in August 2024 at the D2 distances and at the D1 distances.

A.5. Power Analysis Plots



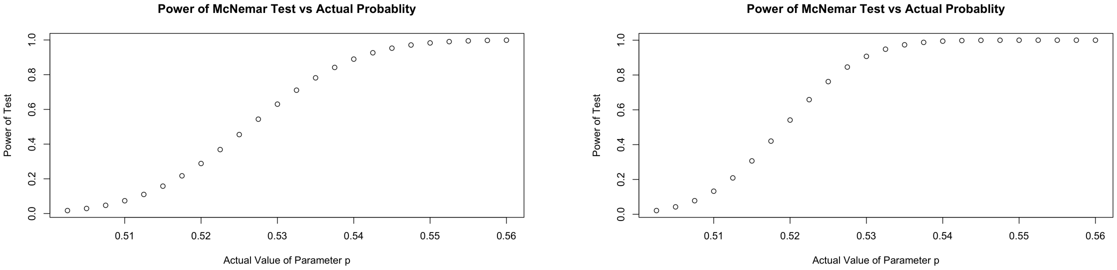
(a) Experiment 1: Power At Several Values For p

(b) Experiment 2: Power At Several Values For p



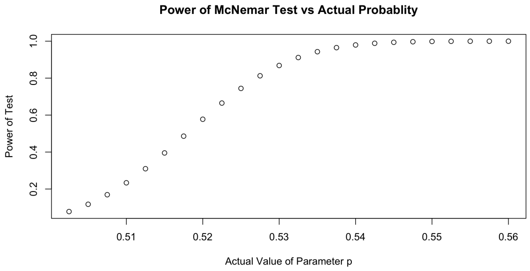
(c) Experiment 3: Power At Several Values For p

(d) Experiment 4: Power At Several Values For p



(e) Experiment 5: Power At Several Values For p

(f) Experiment 6: Power At Several Values For p



(g) Experiment 7: Power At Several Values For p

Figure A-25. Figures A-25a, A-25b, A-25c, A-25d, A-25e, A-25f, and A-25g display the power of the McNemar's test associated with the stated experiment as the value of the p parameter is increased.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Hill, Aaron J	5944	ajhill@sandia.gov
Gonzalez, Efrain Humberto	1421	ehgonza@sandia.gov
Patel, Lekha	1441	lpatel@sandia.gov
Coverdale, Christine A.	5900	cacover@sandia.gov
Kamm, Ryan James	5924	rjkamm@sandia.gov
Horne, Steve	6634	smhorne@sandia.gov
Thoreson, Greg G	6632	ggthore@sandia.gov
Thelen, Paul Mark	6773	pmthele@sandia.gov
Padilla, Eduardo Antonio	6745	eapadil@sandia.gov
Elizondo, Phil	5944	fgelizo@sandia.gov
Technical Library	1911	sanddocs@sandia.gov

Email—External

Name	Company Email Address	Company Name

Hardcopy—Internal

Number of Copies	Name	Org.	Mailstop

Hardcopy—External

Number of Copies	Name(s)	Company Name and Company Mailing Address