Distributed Sensor Network Data Fusion for Nuclear Security Applications

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Ian Robert Stewart
December 2020

Copyright © 2020 by Ian Robert Stewart

All rights reserved.

This is dedicated to my family, friends, and colleagues

who have all helped me along my journey.

To Eirianwen, thank you.

Without your support, encouragement, and near-endless patience,

I would not have been able to complete this work.

Acknowledgements

This work was not possible without the generous support and guidance of so many people. To my family for all of their love, support, and encouragement. To my dearest friends and colleagues in all parts of the country for all of their support, both in the pursuit of my degree and in life. I want to thank my advisor, Dr. Howard Hall, for his dedication, support, and, most of all, his patience over the last few years. I would not have been successful in graduate school without his support.

I also want to thank my committee members from the University of Tennessee-Knoxville, Dr. Jason Hayward and Dr. Lawrence Heilbronn, as their support and feedback make it possible for students to pursue their dreams.

I must also thank Dr. Daniel (Dan) Archer and the Radiation Detection and Imaging group at Oak Ridge National Laboratory. Dan took a chance by supporting my graduate school studies as well as my research activities and I am forever grateful for that opportunity. Dan's mentorship in both my pursuit of a career after graduate school and in life has been invaluable. Additionally, the entire Radiation Detection and Imaging group took me in and gave me the opportunity of a lifetime to learn, grow, and explore a brave new world. Their words of encouragement through this process and the friendships made along the way will not be forgotten. Lastly, I want to show my sincere appreciation to the U.S. Department of Homeland Security Office of Combatting Weapons of Mass Destruction for their financial support during my studies.

Abstract

Detecting, locating, and interdicting radiological materials out of regulatory control or that pose a threat to the general public remains a high national security priority. Conventional methods employed a radiation detection system to detect radiological materials and alert the appropriate authorities for interdiction procedures. More recently, the concept of operations has focused to deploying multiple radiation detection systems to detect and track radiological materials, providing more detail to aid end-users during the interdiction process. The research proposed in this document utilizes data from a distributed sensor network to detect and track radiation anomalies through a network. The primary original contributions of the proposed research focus on creating data fusion techniques between contextual sensors radiation sensors using a Bayesian framework. The Bayesian framework utilizes extracted features from multiple modalities to populate a feature vector to provide more detail regarding the detected anomaly. Utilizing the extracted features from each sensor modality positioned at a node in the network, additional information is inferred, (e.g., source activity, direction, velocity, or time-to-next node), which is sourced to increase downstream sensors' sensitivities. Using the Bayesian framework on sensor data collected from a distributed sensor network on the Oak Ridge National Laboratory reservation, two configurations were analyzed: radiation-only sensor configuration and multimodal sensor configuration. The radiation-only approach is employed to demonstrate how features extracted from the radiation data and metadata can be utilized to detect and track anomalies. The radiation-only sensor configuration also provides a metric of comparison to quantify any improvements additional contextual sensors provide. The multimodal approach expands upon this architecture by fusing the extracted features from the LIDAR and video data to enhance the

ability to track the radioactive materials. Using multiple sensor improved the ability to track the anomalies and correlate signals across the sensor network by as much as 33%. In all anomalous events analyzed within this work, the classification confidence for the *same anomaly* was enhanced by fusing the LIDAR and video data.

Table of Contents

1: Introduction 1

1.1: Organization of Work 2

1.2: Radiation Detection Overview 3

1.3: Relevant Published Literature 20

1.4: Problem Statement 39

2: Distributed Sensor Network Data 41

3: Anomaly Detection and Feature Extraction Methodologies 46

3.1: Radiation Data Analysis 46

3.2: LIDAR Data Analysis 55

3.3: Video Data Analysis 63

4: Distributed Radiation Detection 71

4.1: Energy Spectra Comparison 72

4.2: Region-of-Interest Detection 91

3.3: Analysis of Sensor Array Data 114

4.4: Bayesian Classification Methodology 139

5: Distributed Multimodal Sensor Network 178

5.1: LIDAR Data Fusion 178

5.2: Video Data Fusion 187

5.3: Multimodal Bayesian Classification Methodology 199

6: Conclusion 206

7: Future Work 211

Appendix 230

References 214

Vita 230

List of Tables

Table 1: Classification decision matrix. 8

Table 2: List of radioisotopes with material properties relevant to radiation detection. 15

Table 3: Performance comparison of various object detection and classification models 68

Table 4: Spectral comparison results for L^1 and L^2 norm methods. 92

Table 5: Spectral comparison results for Chi-squared, K-S Test and Cosine Similarity. 92

Table 6: Probability of Detection for multiple detection options on synthetic data. 115

Table 7: Event 1 cosine similarity values for anomalous spectra. 125

Table 8: Event 2 cosine similarity values for anomalous spectra. 132

Table 9: Event 3 cosine similarity values for anomalous spectra. 138

Table 10: Event 1 average Bayesian network output comparison during anomalous event. 201

Table 11: Event 2 average Bayesian network output comparison during anomalous event. 204

Table 12: Event 3 average Bayesian network output comparison during anomalous event. 207

List of Figures

Figure 1: Gamma-ray gross count rate variation with corresponding location in downtown Knoxville, TN. 5

Figure 2: Illustration of classification parameters for two distributions, background (*left*) and anomalous (*right*). [19] 9

Figure 3: Example distributions of an anomaly detection parameter. 11

Figure 4: Example receiver operator characteristic curve for two arbitrary distributions. 11

Figure 5: Example logarithmic receiver operator curve for two arbitrary distributions. 12

Figure 6: Gamma-ray gross count rate variation with corresponding location in downtown Knoxville, TN. 19

Figure 7: Image of the DSN-CC sensor array at LANL. [56] 24

Figure 8: Image of the LLNL Adaptable Radiation Area Monitor system. [58] 26

Figure 9: Passport System Inc. prototype detection devices in support of the DHS-IRSS project. [61] 26

Figure 10: Testbed for DHS-RAIN with radiation detection sensors and video cameras along a roadway. [64] 28

Figure 11: Testbed for DHS-RAIN showing license plate readers position above flowing traffic. [64] 28

Figure 12: D3S RIID from Kromek with gamma and neutron detection capability. [67] 30

Figure 13: Illustration of DHS-MURS sensor configuration. [71] 32

Figure 14: Image of the MURS modular system utilized for urban search operations. [71] 32

Figure 15: Image of the RadMAP system showing various sensors incorporated on the mobile platform. [73] 33

Figure 16: Example RadMAP output from fusing coded-aperture and stitched video data. [73] 35
Figure 17: Example of data processing of image segmentation for background estimation for RadMAP. [74] 37
Figure 18: Velodyne LIDAR and radiation detector configuration at the University of Florida. [76] 37
Figure 19: Example of fusing LIDAR and Radiation data from the University of Florida. [78] . 38
Figure 20: General schematic diagram for processing sensor data utilized in proposed work..... 40
Figure 21: Illustration showing location of nodes in ORNL sensor array. 42
Figure 22: Image of a node (*left*) on the ORNL sensor array along with an image inside the enclosure (*right*). 42
Figure 23: Count rate histogram for nodes in the ORNL sensor array using arbitrary 10-second integration time. 43
Figure 24: Normalized energy spectra for nodes in the ORNL sensor array..... 43
Figure 25: Count rate time series for example anomalous event from ORNL sensor array..... 45
Figure 26: Example of two-window approach for detecting anomalies in time-series data..... 47
Figure 27: Illustration of integration width for mobile source scenario. 50
Figure 28: Optimal integration time for arbitrary mobile radioactive source scenario. [39]..... 52
Figure 29: Receiver operator characteristic curve for gross count anomaly detection algorithm. 54
Figure 30: Histogram of alarm length times for anomalies detected from sensor array..... 54
Figure 31: Example 3D point cloud collected from Velodyne LIDAR unit. [106] 57
Figure 32: Example image of LIDAR point cloud. 59
Figure 33: Overview of LIDAR data processing using ROS and python..... 62
Figure 34: Time series video processing from sensor ORNL using YOLOv3. 69

Figure 35: Example model output from motion event with differing classifications. 69
Figure 36: Example distributions for background and a radioactive source. 74
Figure 37: Source spectrum from two locations. (Top) Example energy spectrum collected during anomalous event at Node01. (Bottom) Example energy spectrum collected during anomalous event at Node02. 74
Figure 38: Background subtracted energy spectra for Node01 (top) and Node02 (bottom). 75
Figure 39: Filtered spectra for Node01. (Top) Filters of various window lengths using a linear model. (Bottom) Filters of various window length using quadratic model. 78
Figure 40: Filtered spectra for Node02. (Top) Filters of various window lengths using a linear model. (Bottom) Filters of various window length using quadratic model. 78
Figure 41: Background-subtracted and filtered spectra for both example anomalous spectra. 79
Figure 42: Normalized and filtered spectra using divide-by-max and sum-to-unity approach. ... 81
Figure 43: Illustration of L^1 and L^2 norm distance measures. 85
Figure 44: Cumulative density functions for Node01 and Node02 anomalous spectra. 87
Figure 45: Illustration of cosine similarity concept comparing two vectors. 89
Figure 46: Background-subtracted spectra with highlighted regions-of-interest using heuristic approach. 95
Figure 47: Filtered spectra with highlighted region-of-interest. 95
Figure 48: Multiple-linear regression detection methodology using region-of-interest approach. 100
Figure 49: Illustration comparing least-squares and ridge regression on arbitrary data. 102
Figure 50: Example illustration of prediction error for least squares and ridge regression methods. 104

Figure 51: Histogram of count rates per bin for a deployed radiation detector. 106
Figure 52: Ridge regression detection methodology using region-of-interest approach. 108
Figure 53: Example ridge trace plot for multiple coefficients using arbitrary data. 108
Figure 54: Ridge trace for radiation detectors deployed at ORNL. 110
Figure 55: Mean squared error for ridge and least squares regression models on each ORNL sensor node. 111
Figure 56: Count rate of synthetic data for comparing detection methods. 112
Figure 57: Receiver operator characteristic curve comparing gross counts and least-squares methods. 112
Figure 58: Illustration of vehicle carrying radioactive material through ORNL sensor array for Events 1 and 2. 116
Figure 59: Illustration of path of a vehicle carrying radioactive material through ORNL sensor array for Event 3. 116
Figure 60: Event 1 time series of count rates observed on the ORNL sensor array. 118
Figure 61: Event 1 alarm and background spectra collected at four nodes. 119
Figure 62: Event 1 region-of-interest from filtered anomalous spectrum at Node06. 120
Figure 63: Event 1 decision metric time-series for remaining nodes. 120
Figure 64: Event 1 anomalous spectra after background-subtracting, filtering, and normalizing. 122
Figure 65: Event 1 cosine similarity time-series. 124
Figure 66: Event 2 time series of count rates observed on the ORNL sensor array. 126
Figure 67: Event 2 region-of-interest from filtered anomalous spectrum at Node06. 128
Figure 68: Event 2 decision metric time-series for remaining nodes. 128

Figure 69: Event 2 alarm and background spectra collected at five nodes. 129
Figure 70: Event 2 anomalous spectra after background-subtracting, filtering, and normalizing. 131
Figure 71: Event 2 cosine similarity time-series. 131
Figure 72: Event 3 time series of count rates observed on the ORNL sensor array. 134
Figure 73: Event 3 region-of-interest from filtered anomalous spectrum at Node04. 135
Figure 74: Event 3 decision metric time-series for remaining nodes. 135
Figure 75: Event 3 alarm and background spectra collected at three nodes. 136
Figure 76: Event 3 anomalous spectra after background-subtracting, filtering, and normalizing. 136
Figure 77: Event 3 cosine similarity time-series. 137
Figure 78: Schematic of single node process for processing radiation sensor data. 142
Figure 79: Schematic of single node process for processing radiation sensor data utilizing previous information. 142
Figure 80: Example network diagram for decision-making process. 146
Figure 81: Energy spectra for Scenario 1 comparing different anomalous spectra from Node04. 152
Figure 82: Energy spectra for Scenario 2 comparing Event 1 and Event 3 anomalies from Node04. 152
Figure 83: Cosine similarity conditional probability for Node01 in Event 1. 154
Figure 84: Example Poisson distribution for arbitrary count rate data. 156
Figure 85: Count rate conditional probabilities for arbitrary count rate data. 156
Figure 86: Expected time-of-arrival conditional probabilities for arbitrary data. 160

Figure 87: Schematic of initial alarm node procedure. 160
Figure 88: Schematic for constructing Bayesian beliefs after initial alarm. 162
Figure 89: Initialization process of prior and conditional probabilities for Bayesian framework. 163
Figure 90: Event 3 posterior probabilities using the Bayesian framework. 164
Figure 91: Event 3 posterior probabilities using the Bayesian framework with the weighted probability approach. 166
Figure 92: Event 1 posterior probabilities using the Bayesian framework. 168
Figure 93: Event 2 posterior probabilities using the Bayesian framework. 170
Figure 94: Event 2 time-series count rate with additional radioactive material present. 172
Figure 95: Event 2 time-series posterior probabilities with additional Am-241 source. 173
Figure 96: Event 2 time-series posterior probabilities with additional Cs-137 source. 173
Figure 97: Event 2 cosine similarity conditional probability values for multiple source scenario. 175
Figure 98: Event 2 time-series posterior probabilities with additional Am-241 source for multiple events. 175
Figure 99: Event 2 time-series posterior probabilities with additional Cs-137 source for multiple events. 176
Figure 100: Geometry of MCNP model utilized in solid angle calculation. 180
Figure 101: Location of point sources (*left*) and associated solid angle (*right*) from MCNP simulations from 2- to 6-meter radius from the sensor. 180
Figure 102: Ordinary least-squares fit to simulated count rate data for source intensity approximation. 180

Figure 103: Updated unnormalized (left) and normalized (right) count rate conditional probabilities for arbitrary count rate data. 182
Figure 104: Example scenario utilized for apparent intensity approximation. 184
Figure 105: Dynamically updated count rate conditional probabilities for example scenario. .. 185
Figure 106: Count rate and OLS fit for example scenario. 186
Figure 107: Example extracted features from LIDAR data with pseudo-classifications. 188
Figure 108: Illustration of data collection between buildings for controlled suppression study. 190
Figure 109: Measured count rate for truck and human surrogate long-dwell controlled suppression study. 192
Figure 110: Example LIDAR output for semi-controlled environment suppression study. 194
Figure 111: Distribution of collected count rates for semi-controlled environment suppression study. 194
Figure 112: Distribution of collected count rates for uncontrolled environment suppression study. 195
Figure 113: Example discrete distribution for arbitrary video classification output. 198
Figure 114: Event 1 time-series of the posterior probabilities using the multimodality Bayesian framework. 200
Figure 115: Event 2 time-series of the posterior probabilities using the multimodality Bayesian framework. 203
Figure 116: Event 3 time-series of the posterior probabilities using the multimodality Bayesian framework. 206
Figure 117: Class list for YOLOv3. 231

1: Introduction

A large amount of resources has been devoted to detecting, locating, and interdicting radiological materials that pose harm to the general public, such as materials for a weapon or dispersal device. Historically, to combat this threat, special nuclear materials (SNM) have been safeguarded and secured by employing detection and material accountancy measures at specific facilities that produce and/or process SNM. As the threat landscape progressed with the emergence of new nuclear powers, the nuclear security regime has focused on improving the effectiveness of deployed sensors positioned away from production and processing facilities, such as utilizing active assay techniques of shipping containers at major seaports or portal monitors positioned at points of ingress [1, 2]. Despite these investments and efforts carried out around the world, over the last two decades, multiple thefts of radioactive materials have occurred. A compiled list of events involving the theft or potential misuse use of nuclear materials can be found in [3, 4, 5] along with additional contextual information.

Unfortunately, several operational parameters decrease the effectiveness of persistently monitoring large-areas or pathways around high-valued assets. Furthermore, in the absence of any ground-breaking technologies that drastically improves sensor performance, the current standard operating procedures utilize detection technologies from decades past, such as thallium-doped sodium iodide (NaI(Tl)) or polyvinyl toluene (PVT) gamma-ray detectors, for static and mobile detection operations [5, 6, 7].

In an effort to counter such scenarios from occurring, current research and development (R&D) efforts focus on utilizing a distributed sensor network (DSN) to improve the overall performance and effectiveness of detecting and tracking radiological material in transit. Using a DSN to integrate available information from multiple detection systems can potentially improve

detection sensitivity, assist alarm adjudication processes to reduce false alarm rates, and track radiological anomalies in the monitored area. To date, R&D efforts have primarily focused on two concept of operations: close-proximity sensor networks or distributed sensor networks. The former has ability to infer parameters regarding the radiological material (e.g., activity, location, or velocity) by observing the signal simultaneously on multiple sensors while the latter does not have the ability for simultaneous observation of the same signal [8, 9]. This work primarily focuses on the utilizing DSN, where the sensors are positioned at locations where simultaneously observing the same radiological material is highly unlikely. This work does not assess large, disparate radiation sensor networks, such as the networks utilized for the Comprehensive Test Ban Treaty Organization or the U.S. Environmental Protection Agency RadNet, as this work primarily focuses on utilizing sensor networks close enough for near-field search operations pertinent to search operations [10, 11, 12].

1.1: Organization of Work

This work is organized to progressively introduce the reader the problem space by first providing a brief overview of concepts that the author deems key to radiation detection in uncontrolled environments, such as an urban area. Following the brief overview, a literature review of relevant published works along with notable sensor network projects within the U.S. is provided. In the second chapter, the dataset utilized that is used to evaluate the methods proposed in this work. Additionally, a brief overview of the equipment, key system parameters, and assumptions that made before the proposed analytical methods can be employed. In the third chapter, the analytical methods used to analyze each specific the data collected from each sensor modality in the dataset. Chapter four introduces a decision-making classification framework using Bayes' theorem to track radiological material that is travelling through an area being

monitored by a sensor array with only radiation detectors. In this chapter, the initial anomaly information is leveraged to obtain multiple data features along with likelihood functions for the purpose of performing a series of hypothesis tests. The hypothesis tests are performed to classify an observed anomaly as either no anomaly, same anomaly, or different anomaly. The work is summarized in the final chapter along with a discussion regarding opportunities for future work.

1.2: Radiation Detection Overview

There exist a variety of methodologies to detect radiation emitted from materials and are generally organized into two categories; active interrogation or passive detection. Active interrogation utilizes other phenomena to stimulate the material to emit radiation, such as using a neutron beam to interact with, or interrogate, a specific material. The material under interrogation has unique interaction properties along with specific emission characteristics which can be subsequently detected in other detectors. The emitted emissions after the initial interrogation can provide informative signals at a higher rate than passive approaches to help identify properties of said material. This approach is ideal for material with small specific activity or high self-shielding [13, 14].

Radiological search operations primarily focus on scenarios wear a potential radiological threat material is traveling in an area open to the general public. Therefore, active interrogation methods are less than ideal as the radiation dose rate drastically increases as large area would be irradiated. For these reasons, passive detection systems are employed for search operations, which are generally smaller and simpler to operate as only a detection system is required compared to an emission source and detection systems for active interrogation [14, 15].

With the goal of sensor network to position static radiation sensors to passively detect and track radioactive materials, several parameters effect observed signal and should be accounted for in an appropriate manner in order to provide meaningful information to an end-user. Several key concepts and relevant parameters to this problem space discussed in the following subsections.

1.2.1: Signal Detection Theory

A primary concept in this problem space is resides in detecting an anomalous signal from a radiation detector. In this use case, the signal observed from the radiation detectors are subject to statistical fluctuations primarily due to radioactive decay being inherently a stochastic process. This noise is exacerbated due primordial radioactivity originating from naturally occurring radioactive material (NORM), such as radioactive isotopes found in soil or physical infrastructure. Three primordial radioisotopes along with their respective daughter isotopes are typically discussed in literature as being dominant; namely, potassium-40, uranium-238, and thorium-232 [16]. The fluctuations observed are referred to as counting statistics, which is used in anomaly detection to quantify the natural fluctuations to approximate the expected signal. One approach to quantifying the noise contribution is to collect data absent of any radioactive sources and constructing a likelihood distribution for the background noise signals [17]. Subsequently, this likelihood distribution can be utilized to detect anomalies by fitting a function to the background noise data. This is conventionally accomplished by fitting a discrete Poisson distribution or, possibly more commonly utilized, a continuous Gaussian distribution and utilizing the underlying distribution's properties to quantify the likelihood of observing a specific sample. This concept can be visualized in Figure 1 for a time-series frequency data for an inserted radioactive source.

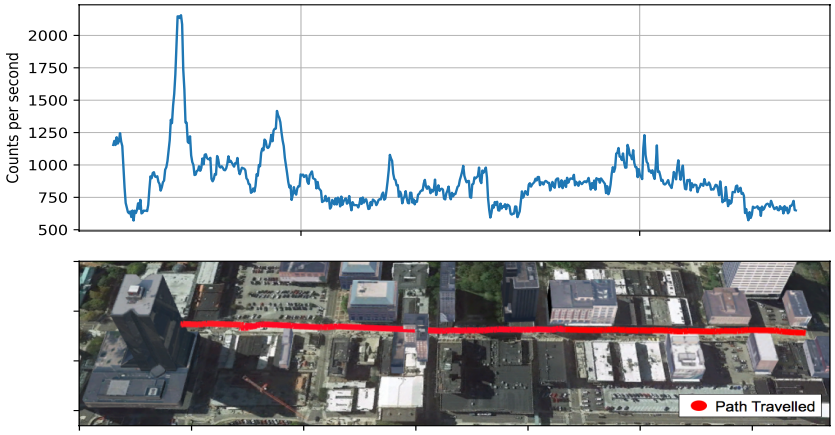


Figure 1: Gamma-ray gross count rate variation with corresponding location in downtown Knoxville, TN.

Classifying observed signals or data as an anomaly is dependent on several factors, such as the width of the background distribution, the separation between the background and anomalous distributions, and the detection methodology output parameter being employed. The detection output parameters are unique for each anomaly detection methodology and can have a complicated distribution (i.e. not a Gaussian distribution). A popular parameter utilized to quantify the detectability of a signal is the signal-to-noise ratio (SNR) that compares an observed new signal with the background noise signal. A relatively common approach to calculating the SNR is defined by the ratio of the mean signal amplitude to the standard deviation of the signal. In this particular case, the SNR can be calculated by the ratio of the radioactive source signal (S) amplitude and the square-root of the source and background signals, $SNR = \frac{S}{\sqrt{S+B}}$. Additionally, in many use cases, the source signal (S) is assumed to be far less tha the background signal (i.e. $S \ll B$), an alternative SNR is utilized, $SNR \approx \frac{S}{\sqrt{B}}$. Using this metric, a large source strength (S) compared to the background ground (B) provides a higher SNR. This results in a greater separation between the source and background distributions and increases the detectability of the source signal when performing an anomaly detection algorithm. The SNR will be discussed in later sections when constructing the anomaly detection methodologies.

For binary classification procedures, a threshold on the detection output parameter is applied to classify collected signals as normal background or anomalous. An approach to selecting the appropriate threshold leverages two types of error, specifically the Type I and Type II error. Type I error is referred to as false-positive (FP) error and represents the error in classifying a signal as anomalous when in fact the signal was a normal background noise. Type II error is also known as false-negative (FN) error and represents the error in classifying a signal as normal background noise when the signal was anomalous. illustrates these relationships

between true and predicted classification along with introducing the true-positive (TP) and true-negative (TN) parameters. In the radiation anomaly detection scenario, the *positive* classification represents that a radioactive material is present, and the *negative* classification means that no radioactive material is present [17, 18]. For additional context, if the two distributions are known for both the background noise and the anomaly, the classification parameters can be visualized as area under the curves, as shown in Figure 2 [19]. The vertical line illustrates the threshold value that can be modified for specific scenarios based on the classification parameters, such false-positive rate.

A common approach to analyzing the data detection methodology and selecting an appropriate threshold value on the output parameter is to plot the detection rate, or true-positive rate, as a function of the false-positive rate. This plot is referred to as a receiver operator characteristic (ROC) curve and helps convey the performance of the anomaly detection methodology being utilized. In real-world scenarios, the distribution of the anomalous signal is likely unknown and a surrogate anomaly, or example radioactive source, is utilized to select a detection threshold. Furthermore, certain operational scenarios can dictate the threshold selection process by determining a threshold value based on a false-alarm rate, as false-alarms increase the effort required to adjudicate the alarm, which can increase costs and decrease overall system efficiency. For example, if the anomaly detection methodology is employed on a radiation detection system for background (i.e. no source present) along with two radioactive sources with unique activities or intensities, a ROC curve can be created based on the output parameter from the detection algorithm, as shown in Figure 3 and Figure 4. The diagonal line labelled *No sensitivity* represents a random guess process, thus ideally the algorithm's values should remain above the diagonal random guess values.

Table 1: Classification decision matrix.

		True Classification	
		Positive	Negative
Predicted Classification	Positive	*True Positive* (TP)	*False Positive* (FP)
	Negative	*False Negative* (FN)	*True Negative* (TN)

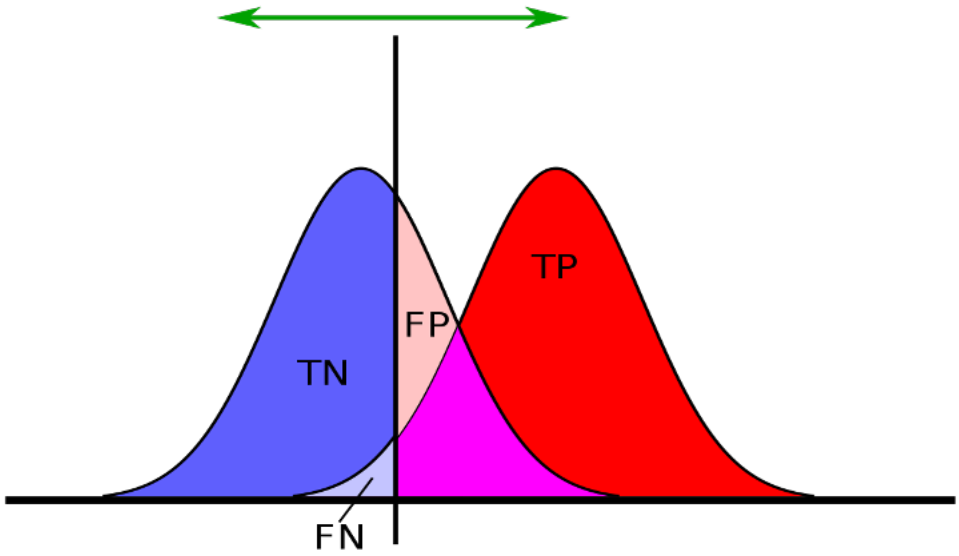


Figure 2: Illustration of classification parameters for two distributions, background (*left*) and anomalous (*right*). [19]

likely unknown and a surrogate anomaly, or example radioactive source, is utilized to select a detection threshold. Furthermore, certain operational scenarios can dictate the threshold selection process by determining a threshold value based on a false-alarm rate, as false-alarms increase the effort required to adjudicate the alarm, which can increase costs and decrease overall system efficiency. For example, if the anomaly detection methodology is employed on a radiation detection system for background (i.e. no source present) along with two radioactive sources with unique activities or intensities, a ROC curve can be created based on the output parameter from the detection algorithm, as shown in Figure 3 and Figure 4. The diagonal line labelled *No sensitivity* represents a random guess process.

In a ROC curve, such as that shown in Figure 4, an optimal classifier would be extremely close to the upper-left hand corner, similar to the *Stronger Source* distribution. The top-right corner of the ROC curve represents a small algorithm output parameter resulting in 100% true-positive rate and 100% false-positive rate. As the anomaly algorithm output parameter is increased, the false-positive rate and the true-positive rate decreases. The ROC curve can also be plotted in log-normal plot, as shown in Figure 5, where the false-positive probability along the X-axis is on a logarithmic scale. This can be useful in visualizing the lower false-positive probability region, which can be used when selecting the appropriate detection threshold.

Uncertainty quantification methods for receiver operator curve are occasionally used to bound the true-positive and false-positive probabilities by providing a degree of belief or error bars [20]. Assuming the classifier is a binomial distribution, multiple methods can be used to quantify the uncertainty, such as normal approximation confidence intervals [21], Wilson confidence intervals [22, 23], or Clopper-Pearson intervals [24, 25]. For instance, if the normal approximation approach is used, the confidence intervals utilize z-scores (z) to provide specific

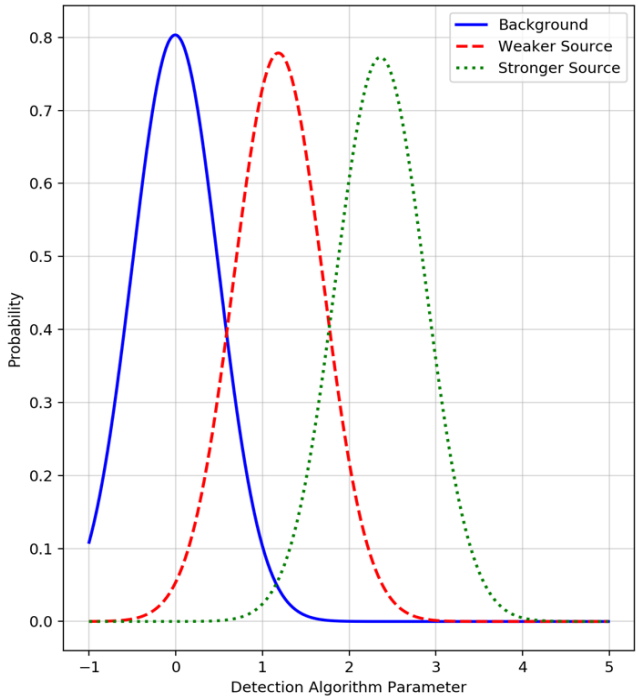


Figure 3: Example distributions of an anomaly detection parameter.

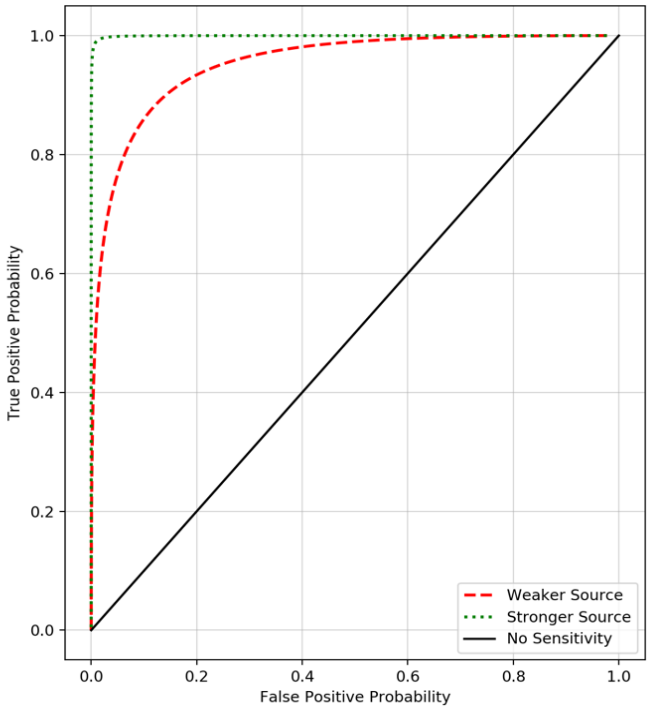


Figure 4: Example receiver operator characteristic curve for two arbitrary distributions.

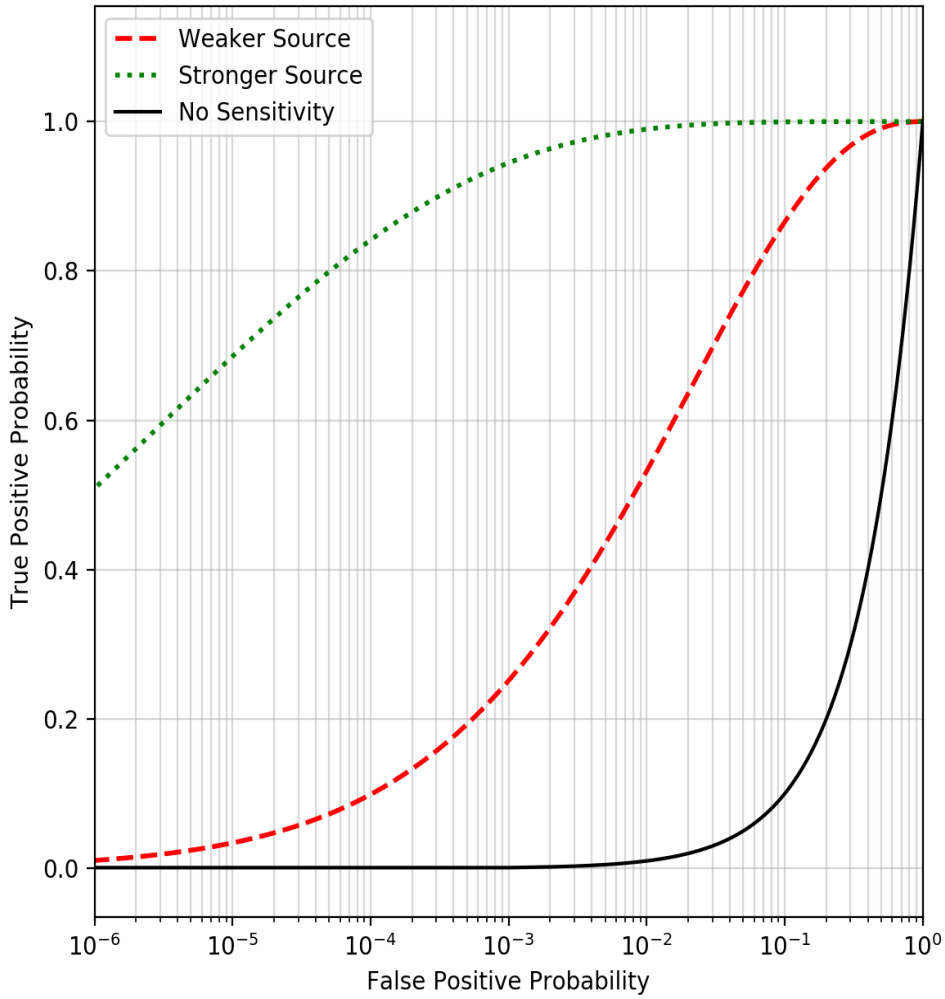


Figure 5: Example logarithmic receiver operator curve for two arbitrary distributions.

bounds to the ROC curve based on the number of success (n_s) and number of failures (n_f) in N number of Bernoulli trials [26] $\frac{n_s}{N} \pm \frac{z}{N\sqrt{N}} \sqrt{n_s n_f}$.

The z-score is obtained using a z-table for a specified confidence level. For example, a 95% confidence results in a target error rate (α) of 5%. This error rate is subsequently used to obtain the quantile for a standard normal distribution in the z-tables resulting in a z-value of 1.96 at 95% confidence. It's worth noting that this method along with constructing the ROC curve is primarily dependent on the particular analyzed data and may not provide identical results for new data. The normal approximation confidence intervals will be utilized in this work to provide some context to the ROC curves for the radiation anomaly detection methodology used to analyze sensor network data.

1.2.1: Type of Radiation Detector

A key consideration in this operational space involves choosing the correct radiation detector to passively detect the radioactive material (e.g., gamma, beta, alpha, or neutron). If the threat radiological material is known beforehand, a rather intuitive and straightforward approach is to utilize a detection system that is highly sensitive in detected the specific radiological material. A list of example radioactive materials is provided in Table 2 along with several material properties [15].

For instance, if the threat radiological material primarily contains cobalt-60 or cesium-137, a gamma-ray detector should be utilized. For special nuclear materials, such as uranium-235 or Plutonium-239, passive detection sensitivity decreases due to a relatively weak specific activity (i.e. number of radiation quanta released by gram of material) coupled with self-shielding effects due to the material's density. If gamma-rays are detected from a gamma-ray emitting radiological source, radioisotope identification is possible as gamma-rays are emitted at

discrete energies following nuclear decay. Additionally, gamma-rays have a relatively large mean free path (i.e. average distance travelled before interacting with a material) in air which provides the capability for standoff detection; ideal for nuclear security scenarios. A primary concern in passively detecting gamma-rays occurs when the radioactive material is shielded by either a cluttered environment between the source and the radiation detector or dense materials are positioned around the material, attenuating and absorbing the radiation emissions [14, 15].

Neutron detection is also a relatively common choice for active assay, passive assay, or radiological search operations due to the particle's penetration power (i.e. ability to pass through dense materials). As neutrons have no charge, neutrons do not interact with the Coulomb force which primarily dominates the energy loss mechanisms as charged particles, such as electrons. Neutrons typically interact with the nucleus of an atom, which can result in the emission of secondary radiations that can be detected (i.e. primarily heavy charged particles) [15]. Additionally, the background, or noise, contribution in neutron detection system is primarily from showers of cosmic rays and remains relatively low, especially compared to gamma-ray detection systems [15, 16, 17]. This low noise contribution results in an increased detection sensitivity and can reduce the over false-alarm rate of detection system. Unfortunately, unlike the discrete emissions for gamma-rays, the energy of the emitted neutrons corresponds to an energy distribution, decreasing the effectiveness in radioisotope identification.

Detection of other radiation emissions, such as the aforementioned beta and alpha radiation, have limited capability in this particular scenario due to the ease in shielding both radiation emissions and the relative short mean free path in air. For these reasons, gamma-ray detectors are typically employed to detect radioactive materials. Worth noting, gamma-ray detectors detect photons emitted from radioactive materials, which also include photons in the x-

Table 2: List of radioisotopes with material properties relevant to radiation detection. [15]

Radioisotope	Approximate Half-life	Specific Activity (Ci/g)	High Energy Alpha Emissions	High Energy Beta Emissions	High Energy Gamma Emissions
Cobalt-60	5.3 years	1100	-	Low Energy	Yes
Cesium-137	30 years	88	-	Low Energy	Yes
Iridium-192	74 days	> 450	-	Yes	Yes
Strontium-90	29 years	140	-	Yes	Low Energy
Americium-241	433 years	3.4	Yes	-	Low Energy
Californium-252	2.7 years	536	Yes	-	Low Energy
Plutonium-238	88 years	17.2	Yes	-	Low Energy
Radium-226	1600 years	1	Yes	-	Low Energy

ray region of the energy spectrum. In this work, the term gamma-ray detector is used as it is a colloquially employed to encompass radiation detectors that detect photons in the energy range relevant in this work; namely, 10 keV to 3000 keV.

1.2.2: Radiation Detector Efficiency

The absolute detection efficiency (ϵ_{abs}) is conventionally expressed as the ratio of signal pulses recorded by the detector system to the number of emitted radiation quanta. This ratio incorporates the intrinsic efficiency (ϵ_{int}) of the detector material to convert the incident radiation into a signal pulse as well as the geometry efficiency (ϵ_{geo}) of the specific detection scenario; $\epsilon_{abs} = \frac{\text{number of pulses recorded}}{\text{number of radiation quanta emitted}} = \epsilon_{int} * \epsilon_{geo}$.

The intrinsic efficiency is primarily dependent on the detection material, the energy of the incident radiation, and the thickness of the material in the direction of the incident radiation. For example, for a gamma-ray detection scenario, an ideal detector would contain a high-Z material with high density whereas a neutron detector generally contains low-Z materials with high interaction cross section for the specific neutron energy [17]. The geometric efficiency, commonly calculated as the solid angle, is of utmost importance in this scenario as the source to the detector distance is generally larger than a laboratory environment, thus decreasing the absolute efficiency. The solid angle (Ω), measured in steradians, is defined as the integral over the face of the detector surface (dA) in the direction of source defined by the angle (α) normal to the detector face at a source-to-detector distance (r), $\Omega = \int_A \frac{\cos(\alpha)}{r^2}\, dA$.

If the source is not a point source, an additional integration must be performed over the volume of the source as well. A relatively common implementation assumes the source can be

modelled as a point source located along the center axis of right circular cylinder with a face area of (a), resulting a simplified calculation, $\Omega = 2\pi\left(1 - \frac{r}{\sqrt{r^2+a^2}}\right)$.

To increase the geometric efficiency, naively increasing the size of the detector material size may not improve the overall sensitivity as the signal from the background radiation could potentially scale at a larger rate than the source signal contribution, decreasing the signal-to-noise ratio (SNR) [28]. As an example, operational solutions to improve the detector sensitivity can be accomplished with longer acquisition times with the material of interest (e.g., static source or decrease traffic speed to increase interaction time) or by decreasing the distance of closest approach (e.g., source-to-detector distance).

The efficiency of a detection system can be estimated by first performing a particle transport model to quantify a relative efficiency as a function of the incident radiation energy. This model will provide an approximation, or rough order of magnitude estimate, for an ideal system, but may not address any loss in efficiencies that can occur in other processes of the detection system (e.g., photomultiplier tube efficiency for scintillators). A fixed experiment with known source emission energies, emission rates, and relative positioning to the detection system can be performed to quantify the total efficiency for the particular detection system.

Often when the experimental data is analyzed, only the photopeak efficiencies are utilized to calculate the efficiencies. This method is less sensitivity to other effects at energies lower than the photopeak that can occur from interactions with materials surrounding the detection system or inherent noise in the system.

1.2.3: Data Acquisition Timing Resolution

It is assumed that the radioactive material of interest is in transit and thus the timing resolution, particularly of the data acquisition system, is a vital parameter. The timing resolution

should at least be set to a value that maximizes the SNR based on a constant background, where an analytical solution as function of source velocity for optimizing the hyperparameter is provided in [29], which will be discussed later in this work. Due to the short interaction time that is inherent to a mobile-source, static-detector scenario, optimal timing resolution allows for the maximal source contribution with limited noise contribution during the collection time.

1.2.4: Spatio-temporal Background Noise

A commonly observed issue in the radiation detection field, especially for mobile search scenarios, is distinguishing background, or noise, from a real source signal. This problem is prominent in urban environments, see Figure 1, as background fluctuations occur due different building compositions, higher probability of nuisance source interaction (e.g., medical isotopes), and variations in angle-of-open-sky (i.e. variations in sky shine contributions) [30, 31, 32]. For a fixed radiation detector, temporal changes are primarily contributed to the dynamic nature of scene surrounding the sensor and from weather-induced signal changes, such as radon washout [33]. These observations are extremely important for the second DSN concept of operations, where spatial differences occur from the unique background or noise contributions and temporal changes may not be observed at all sensor locations.

1.2.5: Other Key Parameters

Two notable parameters that are key in the majority of detection scenarios, but not mentioned here involve gain stabilization and time synchronization. The former involves stabilizing the output signal of the detector in order accurately correlate and compare signals over a period of time, while the latter in concerned with synchronizing the timing across all sensors in a network to cross-correlate collected signals. For more information, the author

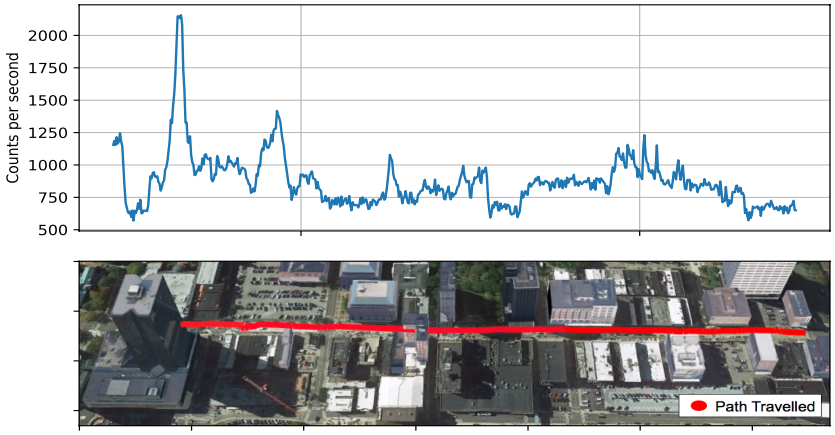


Figure 6: Gamma-ray gross count rate variation with corresponding location in downtown Knoxville, TN.

suggests reviewing several references to understand the central problem and applicable solutions [34, 35, 36].

1.3: Relevant Published Literature

There exists similar research and development efforts expressed in this work in current and past published works from the public and private sector. This section provides a brief review of nuclear detection and security relevant efforts and is divided into two subsections: close-proximity sensor networks and distributed sensor networks. At the end of the section, an additional section is provided that highlights recent relevant research performed utilizing additional contextual sensors for domain awareness and source localization.

1.3.1: Close-proximity Sensor Networks

For a network of sensors positioned in close proximity of one another, often times referred to as a dense network, the primary purpose is to fuse signals from multiple sensors to more confidently detect and locate a radioactive source travelling through the network. Combining the raw sensor data is possible due to the small source-to-detector distances in the network or from a higher relative source activity that provides quantifiable signal contribution to multiple sensors simultaneously. This allows for source parameter inference, such as source activity, location, or velocity, and is primarily accomplished by employing a derivation of inverse-square law, where the isotropic source intensity (S) decreases as a function of distance to the detector (r), not accounting for geometric or intrinsic efficiencies, $I(S,r) = \frac{S}{4\pi r^2}$.

Utilizing this geometric relationship between a source and sensor, two primary methods have been employed to locate a radiological source: analytical solution [37, 38, 39, 40, 41] and Bayesian inference [41, 42, 43, 44, 45, 46, 47]. The majority of published research on this subject attempts to solve for an unknown source position and intensity using known sensor

positions, background count rate, and timing resolution. With this information, the previous equation can be expanded to calculate the expected intensity at each sensor (I_s) for a source activity (S) at a given position (x, y) in the predefined search space, shown in the following equation $I_s(t) = \left[\frac{S}{(x_s-x)^2-(y_s-y)^2} - B\right] * t$. This equation expands the r^2 term in the denominator to match an (x, y) position within the search space and incorporates the background (B) at each sensor location (x_s, y_s) for a given timing resolution (t). Using the observed intensity at each sensor, the equation can be rearranged to solve for the activity at each point in the search space surrounding the sensor network $S_{est}(x, y) = \left(\frac{I_s(t)}{t} - B\right)[(x_s - x)^2 - (y_s - y)^2]$.

With three unknowns (i.e. x, y, and $S_{est}(x, y)$), the analytical approach requires data from at least three sensors observed the source's signal; otherwise, the system is underdetermined. The majority of the published research efforts utilize this relationship to locate, estimate velocity vectors, and track a source through a network. A plethora of unique methods have been proposed to estimate these parameters, which either rely on at least three sensors to observe the phenomena (e.g., genetic algorithms [37, 41], sequential probability testing for triangulation [38]) or attempt to integrate a likelihood function based on simultaneous source observations from multiple sensors (e.g., Bayesian log-likelihood ratio [42], Gaussian kriging [45], Markov Chain Monte Carlo [47]). Unfortunately, these publications do not address the impact on localization for distributed sensors with large distances between each.

1.3.2. Distributed Network Sensing

As shown in the previous section, multiple research efforts have focused on close-proximity detection networks, where multiple, simultaneous observations of radioactive materials can be leveraged to detect, locate, and track materials. Less research in the nuclear security field has been published for distributed sensor networks compared to close-proximity

networks. Distributed sensor networks have conventionally been treated as single detector problems, where each sensor, or radiation detector, operates independently to detect radiation anomalies without leveraging other information within the network (i.e. constituting a *detect-and-wait* scenario scenario). This section highlights notable projects and programs, in chronological order, that utilized distributed sensor networks primarily conducted in the Department of Energy's National Laboratory network [48].

One of the earliest programs that focused on distributed sensor networks in the nuclear security realm was the Detection and Tracking System (DTS) at Lawrence Livermore National Laboratory (LLNL) [49, 50]. The DTS project began in the 1990s, integrating output from fixed and mobile radiation detection systems with information from various other preexisting sensors, such as traffic systems, to detect and track radiation anomalies in an area to guide interdiction operations. Unfortunately, very limited information was published to the general public on the DTS project, therefore the detection and data fusion methodologies and results are not available.

The DTS program efforts at LLNL were continued under the Wide-Area Tracking System (WATS) aimed at fusing a heterogenous sensor array of radiation detectors, photogrammetry, traffic monitoring system data, and various other sensors (e.g., magnetometers, acoustics, or vibration) to a central processing center to process the data for end-users [51, 52]. Similar to the DTS program, the WATS program utilized commercial-off-the-shelf sensors coupled with radiation detectors at fixed locations to monitor an urban area but expanded on the previous efforts under DTS to synthesize and correlate data, both in the temporal and spatial domain. After a radiation anomaly is detected, WATS utilized a Bayesian-based statistical construct to account for expected time-of-arrival at other sensor locations based observed velocity vector to track an anomaly through the network [53, 54]. Identical to the DTS program,

very limited information was published, therefore the detection and data fusion methodologies and results are not available.

Around the same time, the Distributed Sensor Network with Collective Computation (DSN-CC) project at Los Alamos National Laboratory began investigating the use of small, inexpensive, radiation detectors coupled with contextual sensors (i.e. namely, microphones, accelerometers, and magnetometers) in a fixed configuration to track radiation anomalies [55]. The primary deliverables for the program focused on creating a testbed of commercial-off-the-shelf (COTS) sensors and provide preliminary analysis of tradeoffs when combining contextual sensors with radiation signal for tracking anomalies; an image of the constructed testbed is provided in Figure 7 [56].

DSN-CC utilized a simple logical approach for tracking anomalies by calculating the time-of-arrival for subsequent sensors by assuming a constant object velocity and known sensor positions. Moreover, the radiation signal was only processed to detect anomalies when one of the contextual sensors triggered, resulting in a reported lower false-alarm rate. In [57], Nemzek et al. illustrated a method using multiple sensor signals to approximate a vehicle's velocity by employing inverse-square law signal depreciation coupled with a Bayesian construct. Limited data fusion or information sharing occurred between sensors in the network, except the estimated time-of-arrival. Furthermore, even though the attempt was to utilize the sensor network to cover a large area, only close-proximity, linear configurations along a roadway were analyzed in the publications [55, 56, 57].

The Adaptable Radiation Area Monitor (ARAM) at LLNL continued the research efforts from the previous LLNL DTS and WATS projects. The ARAM project constructed and tested portable systems, provided in Figure 8, containing radiation detectors, photogrammetry and



Figure 7: Image of the DSN-CC sensor array at LANL. [56]

global positioning system (GPS) sensors that can be positioned in any configuration to meet an end-user's operational need [58]. The ARAM system utilized list-mode processing of radiation data to help increase signal-to-noise ratio, and thus improve detection sensitivity, along with photographs collected before, during, and after observing an anomalous event. The processed data is subsequently made available on secure web transmission for remote telemetry and end-user viewing. From the available publications and LLNL reports, it is unclear how, or even if, ARAM utilized data other detection systems in the network, or if each system was operated in a stand-alone detection configuration [58].

From approximately 2008 to present, the majority of research in sensor networks for nuclear security has focused on either networks of mobile, small-form factor detection systems or networks fixed, static systems along roadways or at points of ingress. Most notably, the U.S. Department of Homeland Security (DHS) Domestic Nuclear Detection Office (DNDO), now incorporated into the Countering Weapons of Mass Destruction Office, funded two efforts called Intelligent Radiation Sensing System (IRSS) and Radiation Awareness and Interdiction Network (RAIN). The IRSS project was focused on providing equipping State, local, and tribal law enforcement with persistent radiation detection capabilities via a small form-factor radiation sensor with a flexible, wireless network architecture [59]. As one of the grant awardees, Passport Systems Inc. published their efforts in the IRSS program that utilized small, cylindrical NaI(Tl) radiation detectors, image provided in Figure 9, coupled to an Android smartphone device to provide GPS data and communication capabilities to a base station [60, 61].

During experimentation and testing phases of the DHS-IRSS project, these detector systems were deployed in fixed locations as well as on mobile system to survey a specific urban area. Each detector relayed radiation data (i.e. raw counts) to create a grid-based, radiation



Figure 8: Image of the LLNL Adaptable Radiation Area Monitor system. [58]



Figure 9: Passport System Inc. prototype detection devices in support of the DHS-IRSS project. [61]

background map in order to obtain a baseline of the area for which to employ an anomaly detection algorithm. Upon initial detection of an anomaly on a single system, other adjacent systems in the specific region are sampled to localize the source based on the collected raw counts. In the publications, the developed algorithms required multiple observations of a source, either by dense network positioning or relatively strong radioactive source strength [62]. Similar to the close-proximity network methodologies expressed by Cunningham [53] and Lanman [55], Bayesian inference approach was utilized by the DHS-IRSS vendors to incorporate the detector systems in the region to localize and track radioactive materials [60, 61]. Due to the requirement for simultaneous observations on multiple detector systems, the published work from this program does not address the issue of detecting and tracking anomalies in a truly distributed sensor network.

The DHS-RAIN project investigated the benefits of fusing radiation detection with contextual sensors, namely video cameras and license plate readers, for monitoring roadways for radioactive threat materials with minimal impact to free-flowing traffic [63, 64]. Under the DHS-RAIN project, a collaboration between Passport Systems, Inc. and Charles River Analytics, Inc. created an experimental testbed of COTS sensors to collect data in a controlled environment, as shown in Figure 10 and Figure 11 [64]. This collaborative effort focused on a feature-level data fusion, similar to the proposed research in this report, by extracting features from each modality and providing that information to an end-user for data fusion. The video cameras were utilized to track vehicles and obtain coherent vehicle tracks in order to attribute an observed radiation signal to a vehicle in scene. Utilizing an inverse-square law approximation, the vehicle that most closely meets the expected fit to raw radiation counts with the maximum likelihood estimate is selected to be the source-carrying vehicle [62]. Of note, there existed no data fusion architecture



Figure 10: Testbed for DHS-RAIN with radiation detection sensors and video cameras along a roadway. [64]



Figure 11: Testbed for DHS-RAIN showing license plate readers position above flowing traffic. [64]

between each of the deployed radiation sensors nor was there an approach to incorporate prior knowledge from previous detection events [62, 64]. This is key to highlight as this proposed research will attempt to utilize both of these factors.

In 2014, the Defense Advanced Research Project Agency (DARPA) SIGMA program aimed to significantly improve the overall cost of deploying a city-scale, mobile, small form-factor spectroscopic systems with robust networking capability [65]. This program utilized each detection system to continuously survey an area and send information to a central node to process the spectral data and update a radiation map of the region. Deploying detection systems made by Kromek Group PLC, such as the D3S system, the program has successfully deployed the smartphone-size systems to first responders in large metropolitan areas with more than 1000 sensors surveying an area simultaneously [66, 67]. Unfortunately, limited information is publicly available regarding the algorithms and results for anomaly detection and fusion of the data. The primary reported successes highlight the capability to successfully network a large number of sensors (i.e. greater than 1000) and stream the data to a central node for processing [68].

1.3.3: Additional Relevant Works

This section highlights other relevant research efforts to the proposed work in this document, primarily focusing on the use of orthogonal sensor modalities for increased situational awareness. The research efforts presented here are not sensor networks, thus were not included in the previous two subsections. There exist three notable projects that utilize additional sensor modalities in conjunction with radiation sensors:

1. Mobile Urban Radiation Search (MURS) project [69, 70, 71],
2. Radiological Multi-sensor Analysis Platform (RadMAP) project [72, 73], and



Figure 12: D3S RIID from Kromek with gamma and neutron detection capability. [67]

3. Radiological Source Detection and Tracking Based on Multi-Sensor Data Fusion project [76, 77, 78].

The MURS project, funded under U.S. DHS-DNDO at Lawrence Berkeley National Laboratory (LLBL), constructed and tested a modular designed detection system to be used within vehicle platform during radiological search operations [69, 70]. The system utilized six NaI(Tl) radiation detectors positioned in a cylindrical configuration to localize a radioactive material by comparing count rates in each detector as well as two ^{6}LiF/ZnS neutron detectors to be positioned along both sides of the vehicle. Upon detection of a radioactive source, the MURS system begins capturing information from the non-radiological sensors, namely video cameras, infrared cameras, and LIDAR units, to provide additional information to the end-user for attribution. See Figure 13 and Figure 14 showing the MURS configuration [71]. The primary data fusion between radiation and contextual sensors that has been reported or published on the MURS project occurs by providing images to the end-users of the region estimated from the cylindrical NaI(Tl) module. Limited information is available on truly fusing the data streams or extracted features from the data streams to either infer anomaly properties or provide additional context to an end-user [71].

The RadMAP system, a DHS-DNDO funded effort at LLBL similar to the DHS-MURS project, outfitted a large truck platform with a multitude of sensors to study emerging concepts in mobile search scenarios [72]. The system, cutaway view provided in Figure 15, contained a 10 x 10 array of NaI(Tl) detectors with a coded aperture mask made of lead for imaging, two arrays of high-purity germanium (HPGe) detectors, an array of liquid scintillator neutron detectors, two Velodyne LIDAR units, two video cameras, hyperspectral imaging systems, GPS units, and an

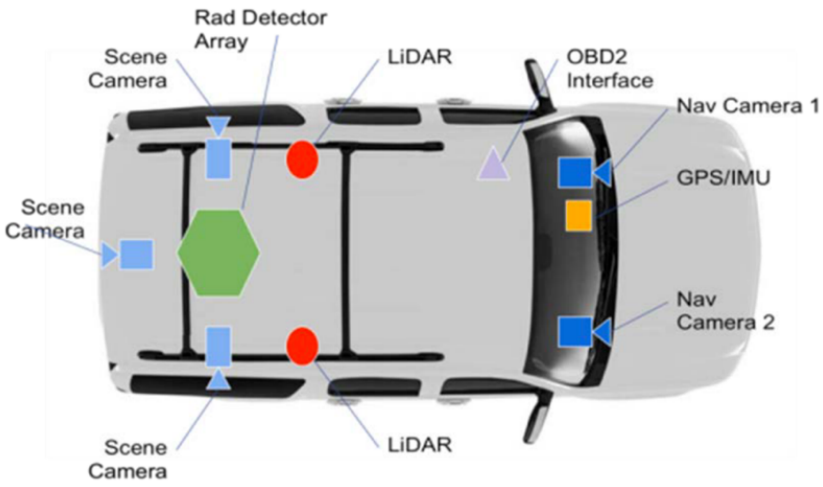


Figure 13: Illustration of DHS-MURS sensor configuration. [71]

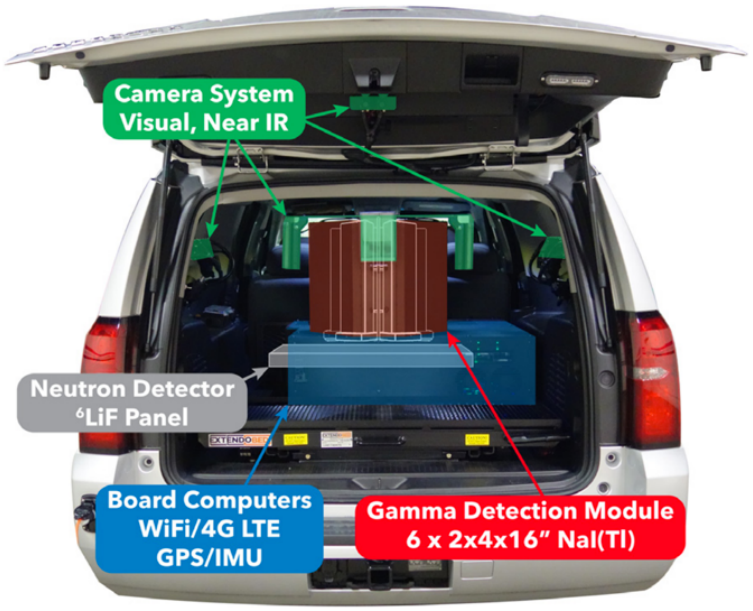


Figure 14: Image of the MURS modular system utilized for urban search operations. [71]

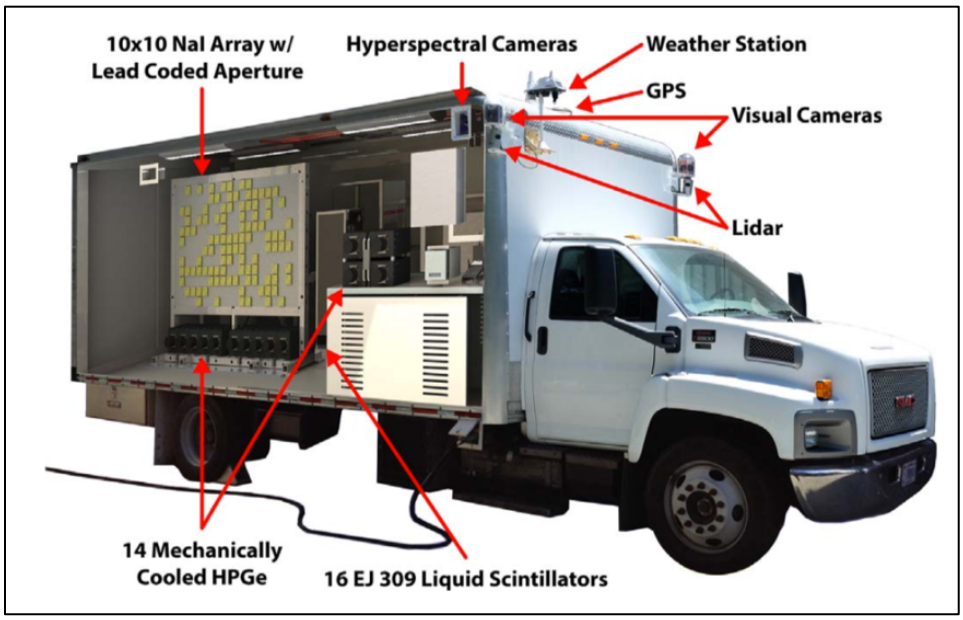


Figure 15: Image of the RadMAP system showing various sensors incorporated on the mobile platform. [73]

onboard weather station [73]. Utilizing a large array of radiation detectors for imaging coupled with the high-resolution HPGe systems provides the ability to detect, locate, and identify radiological anomalies during mobile search operations. The RadMAP project research efforts has utilized the data from the contextual sensors for two primary uses: (1) localize a radiological source in 3D space and (2) online radiation background estimation [72, 73]. Using the output from the NaI(Tl) coded mask array, a radiological source position is approximated using the angle and depth calculations over a series of time intervals. In a post-process manner, these estimates are subsequent used to project the position on the stitched video data and LIDAR data; an image of this process is provided in Figure 16 [73]. Utilizing a large array of radiation detectors for imaging coupled with the high-resolution HPGe systems provides the ability to detect, locate, and identify radiological anomalies during mobile search operations. The RadMAP project research efforts has utilized the data from the contextual sensors for two primary uses: (1) localize a radiological source in 3D space and (2) online radiation background estimation [72, 73]. Using the output from the NaI(Tl) coded mask array, a radiological source position is approximated using the angle and depth calculations over a series of time intervals. In a post-process manner, these estimates are subsequent used to project the position on the stitched video data and LIDAR data; an image of this process is provided in Figure 16 [73].

The second data fusion technique was for radiological background estimation during search operations. To estimate the expected background contribution, the video data is segmented using edges calculated within the video field-of-view (FOV) and is subsequently classified using a random forest classifier into a class, or material, category (i.e. grass, building, asphalt, or vehicle), as illustrated in Figure 17 [74]. Each classification has a predetermined spectral shape and activity [75]. A solid angle is then calculated between the segmented regions

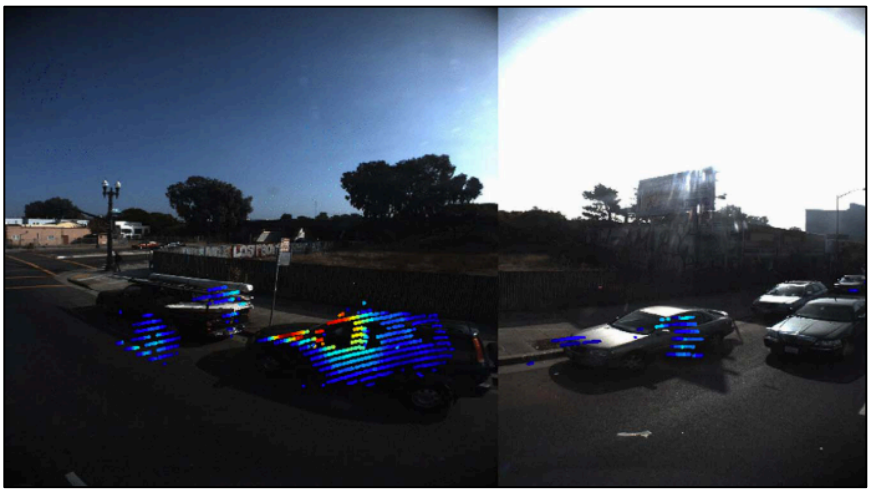


Figure 16: Example RadMAP output from fusing coded-aperture and stitched video data. [73]

and the effective area of the detection system. A linear attenuation is applied for specific photon energies within the spectral information for each class category to estimate the expected background contribution by the segmented region. The results provided in [74] provide reasonably accurate results for estimating background contributions on the detection system, but can vary rather wildly if the preset spectral profile or activity of the material is not accurately known. Furthermore, image segmentation is highly dependent on edges or colors in the image, thus can provided less than ideal results if, for example, a material is different than the expected color. Important to note in regard to the discussed RadMAP project, a rather common problem with fusing video or LIDAR data, especially 3D pointclouds, is the computational power required to process each data stream. This is only compounded in the mobile search operation, when the reference frame is constantly changing within a fully dynamic scene.

The third relevant research effort highlighted in this subsection is a DHS-DNDO Academic Research Initiative project conducted at the University of Florida (UF) that coupled a Velodyne LIDAR unit with a radiation detector to localize a mobile radiological source, as shown in Figure 18 [76]. The research efforts utilize a down sampled 2D pointcloud from the LIDAR and track objects in the LIDAR FOV using a Kalman filter. The positions of the objects tracked are utilized along with the raw count rate to calculate a correlation metric based on the inverse-square law. The cosine distance and the Pearson correlation methods were employed as the correlation metric between the position and observed count rate, where the object with the highest correlation was chosen as the source carrying object, as shown in Figure 19 [76, 77, 78]. The work continues to provide a method for multiple source being present in the scene by expanding the single source search to a source search equal to the number of detected objects [78]. This method allows for continuous searching for source when an object is detected in the

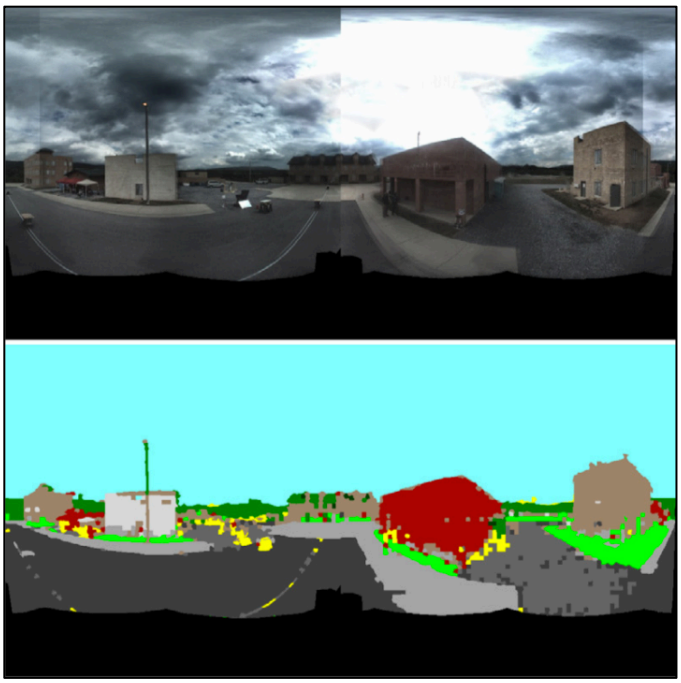


Figure 17: Example of data processing of image segmentation for background estimation for RadMAP. [74]



Figure 18: Velodyne LIDAR and radiation detector configuration at the University of Florida. [76]

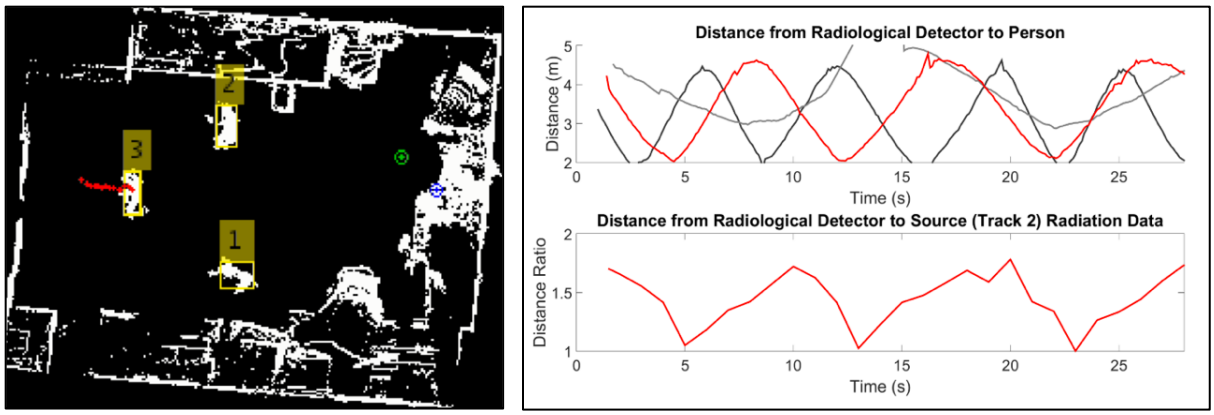


Figure 19: Example of fusing LIDAR and Radiation data from the University of Florida. [78]
(*Left*) 2D LIDAR pointcloud showing objects tracked in yellow boxes. (*Right*) Observed distances of objects from LIDAR and inverse-square ratio of highest probable object.

pointcloud. The UF work provides a relevant example of feature-level data fusion between contextual sensor (e.g., LIDAR) and a single radiation sensor at a fixed position. The efforts performed do not to extract other features, such as velocity or object classification, which this work believes to be useful to tracking radiation anomalies when fusing the LIDAR data with radiation data. Furthermore, since the UF efforts focused on single, fixed locations; there is no method to utilize the extract features at other locations or nodes. Both of these comments will be addressed in the proposed research in the document.

1.4: Problem Statement

This work expands upon the previous approaches utilized in distributed sensor networks for tracking mobile radioactive materials. This is accomplished by utilizing data from the initial detected anomaly to increase detection sensitivity and correlate subsequent signals at other sensors with the initial anomalous data. Multiple features are extracted from the initial anomalous data and are used in a Bayesian-based classification method to track radioactive materials in the area. Additional, collocated sensors (i.e. namely, video data and LIDAR point clouds) are analyzed in order to improve the classification method, improve detection sensitivity, and reduce false-alarm rates. For context, a high-level schematic diagram is provided in Figure 20 showing the general process of combining the primary sensor (i.e. radiation detector) and secondary sensors (i.e. LIDAR or video). In the spirit of full disclosure, the individual methods utilized to analyze each respective data stream may not be by itself novel, but the fusion of the data streams coupled with the created decision-making frameworks have yet to fully analyzed in literature. For this reason, the research efforts presented in this work provide a novel approach to feature-level data fusion for improved detection and tracking of radiological threats travelling through a distributed sensor network.

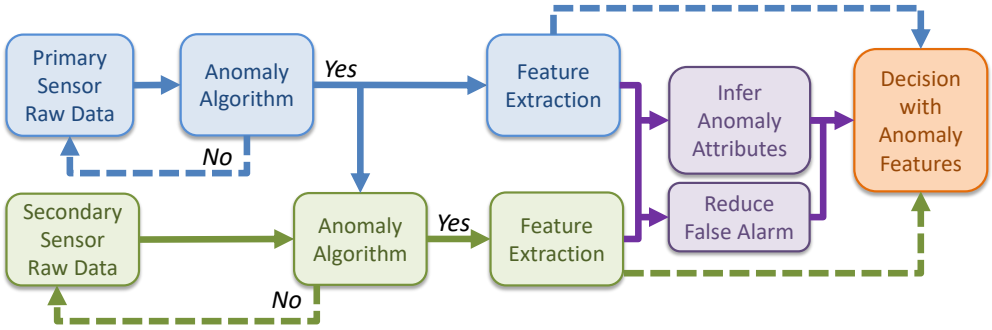


Figure 20: General schematic diagram for processing sensor data utilized in proposed work.

2: Distributed Sensor Network Data

The data analyzed in this work was collected from a sensor network deployed on the campus of Oak Ridge National Laboratory. The sensor array is positioned around facilities that handle radiological materials and observes rather unique signals during facility operations, as shown in Figure 21. The enclosures, denoted to as nodes, are referred throughout this work by the number specified in Figure 21; specifically, Node01, Node04, Node06, Node10, Node11, and Node12. The sensors are located within temperature controlled, weather-proof enclosures and utilize a network time protocol for time synchronization for the sensors across the network. With six nodes operational, each node, see Figure 22 for an image of a node on the ORNL sensor array, in the network contains a 2” × 4” × 16” NaI(Tl) radiation detector, 3” × 3” NaI(Tl) cylindrical radiation detector, Davis Instruments weather station, GPS modules, Velodyne VLP-16 Puck LIDAR sensor, and a single-board computer operating the data acquisition system (DAQ) [79, 80, 81]. Additionally, a network of video cameras is positioned in the area and is time synchronized to provide additional contextual information during the data analysis [82].

The primary sensors utilized in this work include the 2” × 4” × 16” NaI(Tl) radiation detector, Velodyne VLP-16 Puck LIDAR sensor, and the video cameras. The radiation detectors are considered the primary sensors that initiate an anomalous event with the LIDAR and video cameras referred to as secondary sensors that provide additional, or contextual, information during anomalous events. The radiation signal observed at each location varies from node to node, both in signal intensity and spectrally, as shown in Figure 23 and Figure 24. The variations in the observed signal is attributed to the unique background signals originating from naturally occurring radioactive materials observed at different geospatial locations.

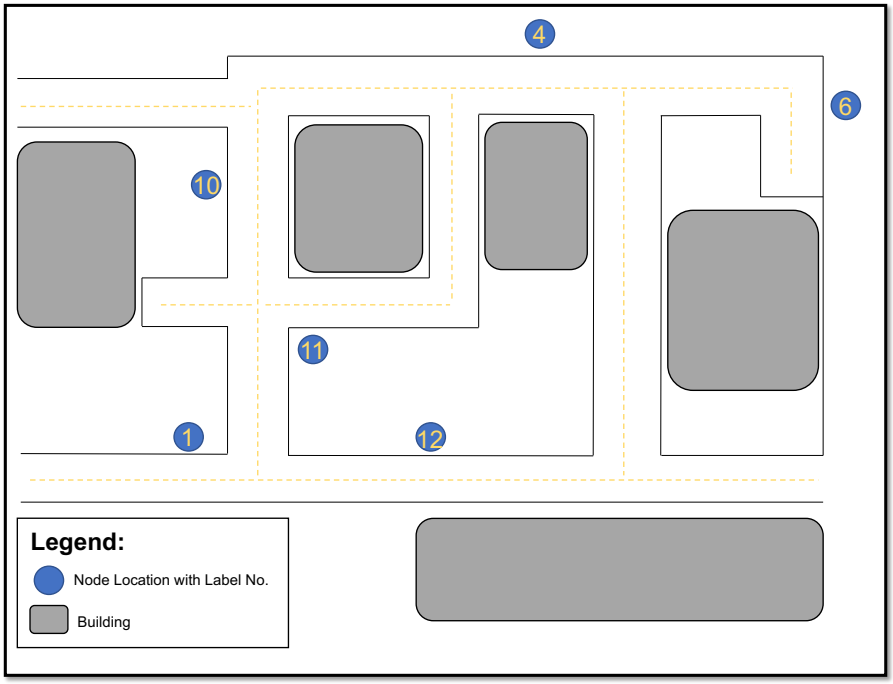


Figure 21: Illustration showing location of nodes in ORNL sensor array.

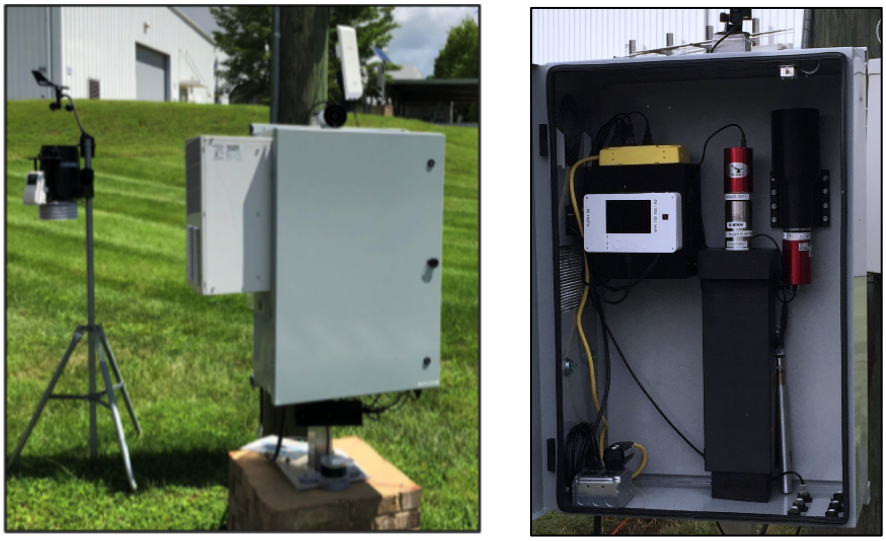


Figure 22: Image of a node (*left*) on the ORNL sensor array along with an image inside the enclosure (*right*).

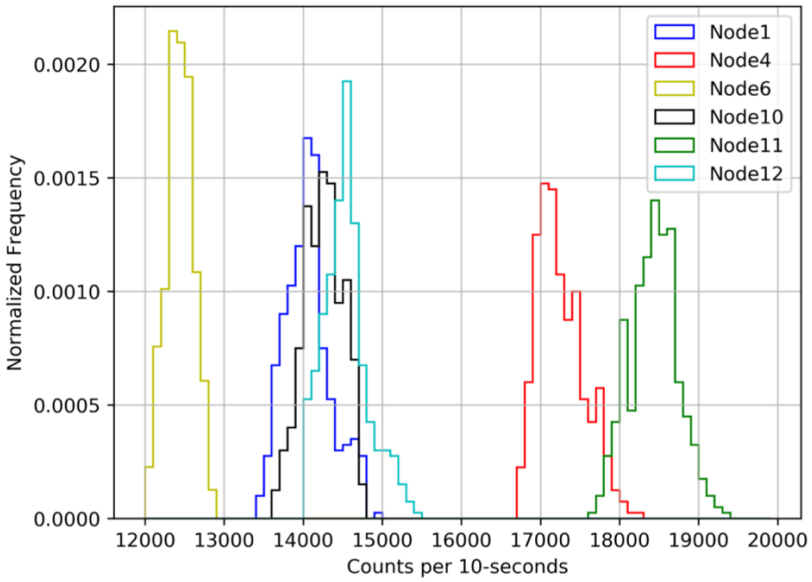


Figure 23: Count rate histogram for nodes in the ORNL sensor array using arbitrary 10-second integration time.

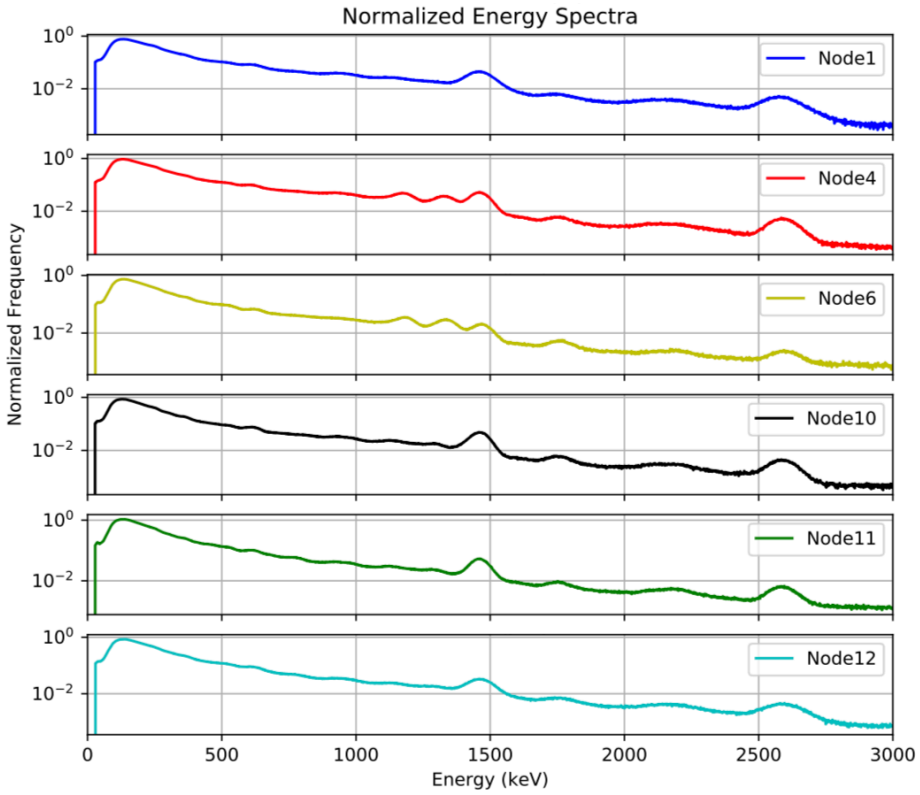


Figure 24: Normalized energy spectra for nodes in the ORNL sensor array.

The radiation data analyzed in this work are gain-stabilized by tracking the gamma-ray emission from K-40 at approximately 1460 keV. This allows for a more accurate comparison between spectra collect at different times on the same node as well as comparing spectra collect from different nodes, as the amplification gain in the photomultiplier tubes and associated detector electronics can shift over time [83, 84]. An energy calibration was initially performed using Eu-152 and Am-241 radioactive sources to provide multiple photopeaks for fitting a calibration curve for a 3 keV binning structure from 0.0 to 3000.0 keV. The data from the DAQ are collected 0.1 second time resolution but are summed to 1 second resolution in this work, as shown by the frequency in the example scenario provided in Figure 25 along with the corresponding images observed from the video data. Analysis and feature extraction for each data stream is described in the following chapter.

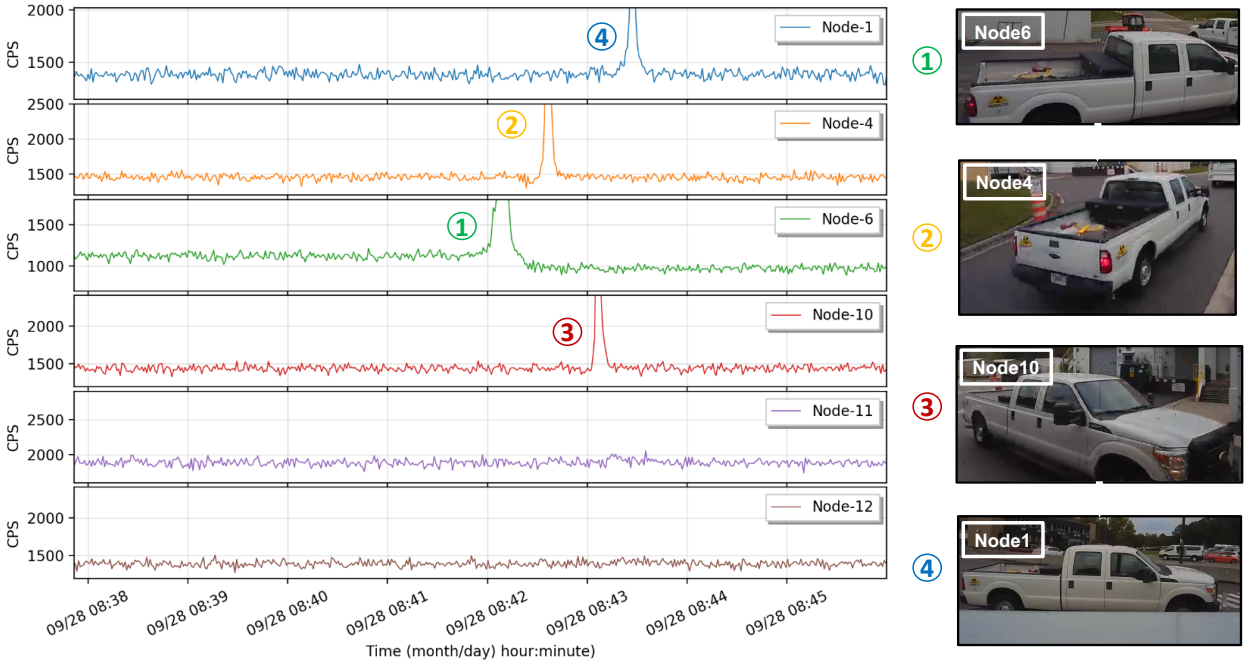


Figure 25: Count rate time series for example anomalous event from ORNL sensor array.

3: Anomaly Detection and Feature Extraction Methodologies

This chapter introduces and describes the methods utilized to analyze three data streams; radiation data, LIDAR point clouds, and video data. The analysis methods are utilized to extract specific features from the respective data stream to construct a feature-level data fusion, discussed on later sections of this work.

3.1: Radiation Data Analysis

As the primary anomaly detection sensor in this work, the data collected from the 2" × 4" × 16" NaI(Tl) radiation detector require a robust analysis method as each sensor has unique signal features; namely, intensity and spectral shape. The primary goal in analyzing the radiation data is to detect anomalies and outliers in the signal with a high true-positive rate (TPR) and low false-positive rate (FPR). Time-series anomaly detection is utilized in numerous of domains with a plethora of published research in recent years [85, 86, 87]. Furthermore, there exists a multitude of algorithms specifically for the use of detecting anomalous signals in real-world radiation data [88, 89, 90, 91].

In this work, a conventional gross count rate approach is utilized to detect anomalies in the total count rate across the entire energy range (i.e. 0 to 3000 keV). This approach is referred to as the gross counts detection method and utilizes two rolling-windows to compare the expected count rate from the background window with the count rate in the foreground window. This is concept, shown in Figure 26, is a relatively simple approach to detect anomalies in time-series data.

The background and foreground windows are separated by a *dead zone* of a few seconds, in this case. This allows for improved detectability by limiting the chance a slowly increasing

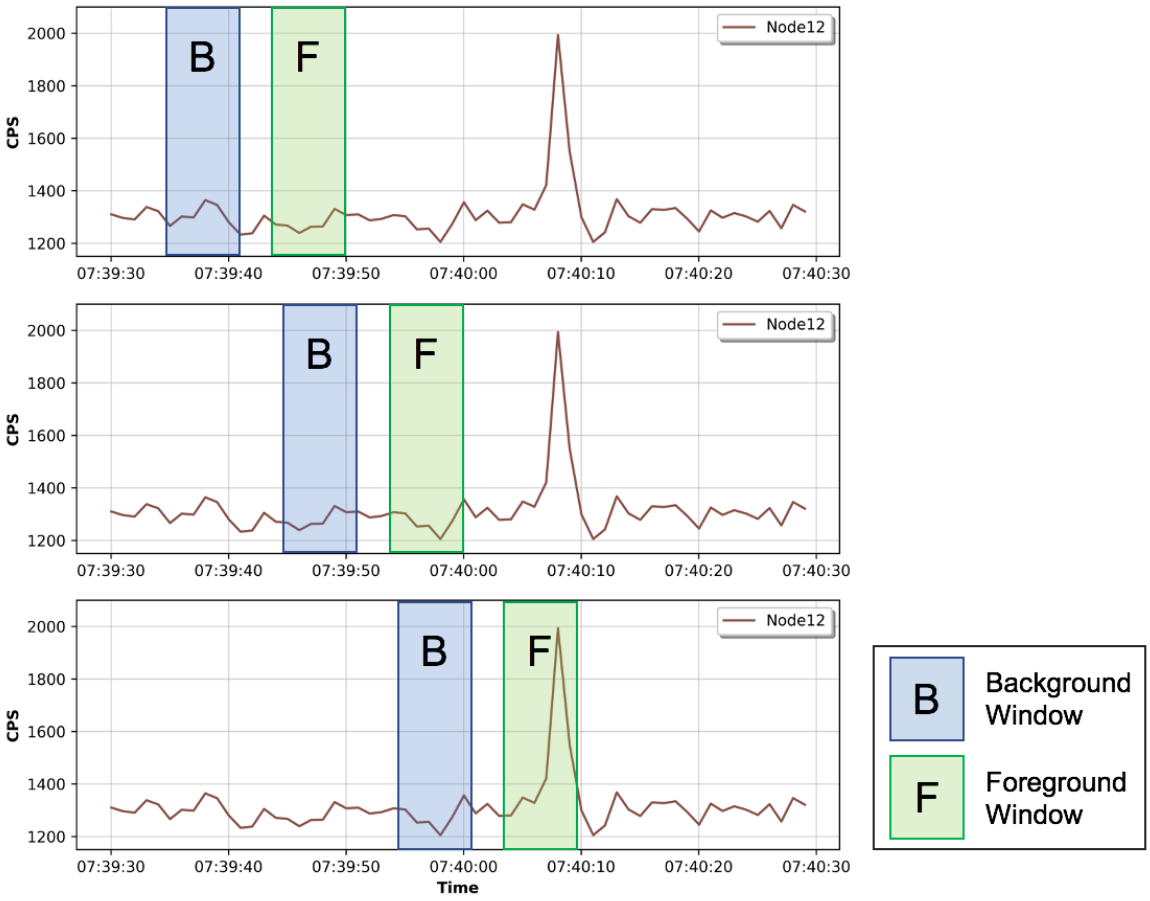


Figure 26: Example of two-window approach for detecting anomalies in time-series data.

anomaly or a low-frequency anomaly from increasing the expected count rate in the background window prior classifying the anomaly in the foreground.

The gross count rate data within each window is assumed to be a sample from a Gaussian distribution, therefore the mean gross count rate (μ) within the window is obtained by dividing the summation of the total counts in the window by the total window live time (t). A decision metric (D) is obtained by calculating the difference between the acquired means for each window and dividing by the difference by the square-root of the background mean (i.e. standard deviation as the sample is assumed Gaussian distributed). This decision metric is provided in the following using the respective parameters for the background (B) and foreground (F) windows shown in the subscript, $D = \frac{\left(\frac{\mu_F}{t_F} - \frac{\mu_B}{t_B}\right)}{\sqrt{\frac{\mu_B}{t_B}}}$.

Two key hyperparameters are required for calculating the decision metric for this approach; size of the rolling-windows and the detection threshold. The detection threshold is dependent on the size of both the background and foreground rolling-windows, therefore it is address first. If the windows are too small, especially for the background, the sample mean can vary greatly from the population or true mean count rate due to high variability in a small sample size. Furthermore, the standard error depends directly on the sample size [17, 92]. Furthermore, there are differing opinions when determining the minimum sample size required to justify the assumption that data is Gaussian distributed governed by the central limit theorem [92, 93, 94]. The central limit theorem states that the sample mean with a large sample size of independent random variables is approximately Gaussian distributed [95]. In this work, a background sample size of 30 it utilized, which is in itself rather controversial and a window size greater than 30 can provide a more accurate representation of the true mean count rate of the background.

The foreground window size is unique in that the window-size should be optimized to maximize the SNR (recall from page 6) by limiting the background noise contribution to the mean of the window. In addition, the window size should be large enough to remain robust to single-instance anomalies that can occur inherent to most statistical processes. To help bound this issue, previously published efforts are utilized. Aucott et al. [29, 96] and Nemzek et al. [39] arrive at similar solutions from different approaches to calculating the optimal window-size during mobile search operations of a static, stationary radioactive material. In order to maximize the SNR, the original SNR is initially expanded to incorporate the full source and background terms assuming a fixed offset distance (h). Figure 27 illustrates the scenario at hand, which was recreated from [29] to provide context for the reader. The count rate observed from the radioactive material can be generally thought to follow an inverse-square function, aside from the solid angle, attenuation, or detector efficiencies parameters which also govern the observed count rate.

The integration with respect to time in both [29, 96] and [39] is changed to the spatial domain, referred to as the interaction length in [39], and calculated by rearranging the conventional velocity formula to obtain the time parameter (dt); that is, $v = \frac{dx}{dt}$ to $dt = \frac{dx}{v}$. If the interaction length is specified at the center of the detection (i.e. $x = 0$), the integrals can be performed over the range $[-x, +x]$. Assuming a constant standoff distance (h) and a constant background contribution, rearranging the velocity formula and modifying the dt term, the SNR can be expanded to incorporate multiple parameters, $S = \int_{-x}^{+x} \frac{S_0}{x(t)^2+h^2}\,dt = \int_{-x}^{+x} \frac{S_0}{x(t)^2+h^2}\frac{dx}{v} = \frac{2\,S_0}{v\,h}\arctan\left(\frac{x}{h}\right)$ and $B = \int_{-x}^{+x} B\,dt = \frac{2\,B\,x}{v}$. Solving for SNR, $SNR = \frac{S}{\sqrt{S+B}} \approx \frac{S}{\sqrt{B}} = \frac{\frac{2\,S_0}{v\,h}\arctan\left(\frac{x}{h}\right)}{\left(\frac{2\,B\,x_0}{v}\right)^{0.5}}$.

Using the expanded the SNR formula, the optimal spatial window is calculated at the peak of the derivative with respect to (x) acquired by setting the derivative equal to zero, the particle

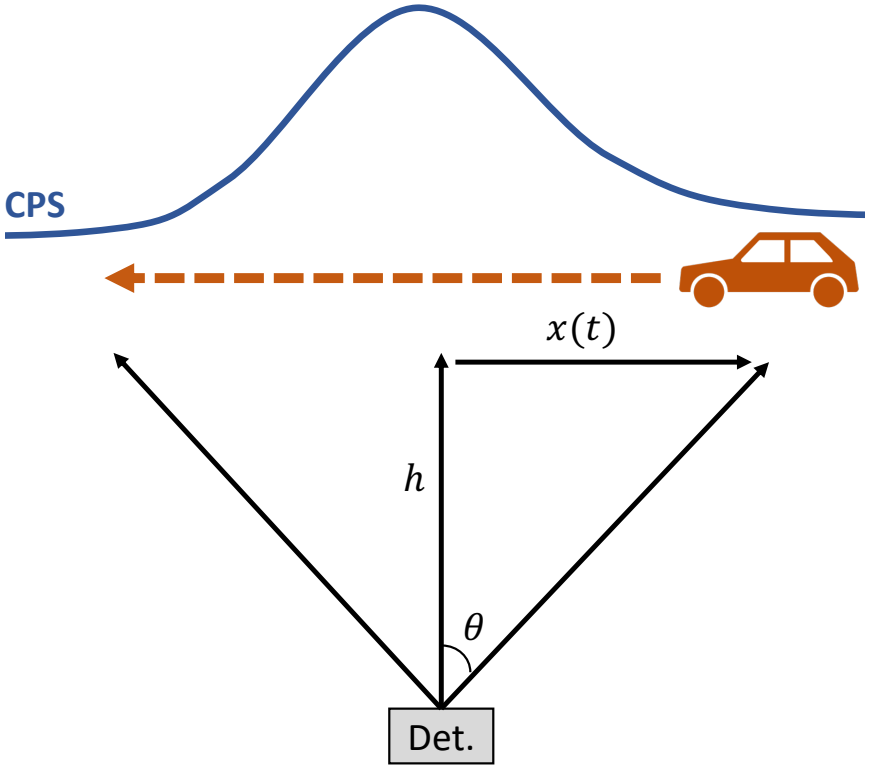


Figure 27: Illustration of integration width for mobile source scenario.

deriviate can be obtained, $\frac{\partial SNR}{\partial x} = \left(\frac{\frac{2 S_0}{v y} \arctan\left(\frac{x}{h}\right)}{\left(\frac{2 B x_0}{v}\right)^{0.5}} \right) \frac{1}{\partial x} = 0$. The numerator of the partial derivate is subsequently set to equal zero $2xy - (x^2 + y^2) \arctan\left(\frac{x}{y}\right) = 0$. Aucott et al. [29, 96] solves the quadratic numerically to arrive at the following relationship, $x = 1.39\, y$.

This allows for an optimal integration window if the expected vehicle offset and speed is known. In [29, 96] and [39], the authors note that even with an optimal window calculating, the addition of the operational parameters (e.g., attenuation through air or solid angle changes) can greatly alter the relationship. For example, Nemzek et al. calculate the optimal integration time, referred to as timing resolution in this work, as a function a velocity of the radioactive materials at a given distance of closest approach and source activity (i.e. 20 meters and 10 milliCurie). The results are provided in Figure 28 and additionally provides an optimal integration time that maximizes the SNR for the range of velocities analyzed shown by the dashed line [39].

As the velocity increases, the SNR decreases, which is expected as the time of interaction decreases. Figure 28 additionally provides an optimal integration time that maximizes the SNR for the range of velocities for the specific scenario analyzed shown by the dashed line [39]. A primary concern when performing the above approaches is the calculated integration window is optimized for a specific scenario and can decrease the robustness as the integration window may be suboptimal for different scenarios [29, 39, 96]. Having speed and distances estimates could allow window optimization to assist in calculating an optimal foreground window for the anomaly detection methodology utilized in this work. The monitored area within the ORNL sensor array typically observes pedestrians and vehicular traffic ranging from 1 to 4 meters distance of closest approach. The velocity varies between pedestrians and vehicular traffic, but range from 4 to 35 miles per hours, or approximately 2 to 15 meters per second. If the primary

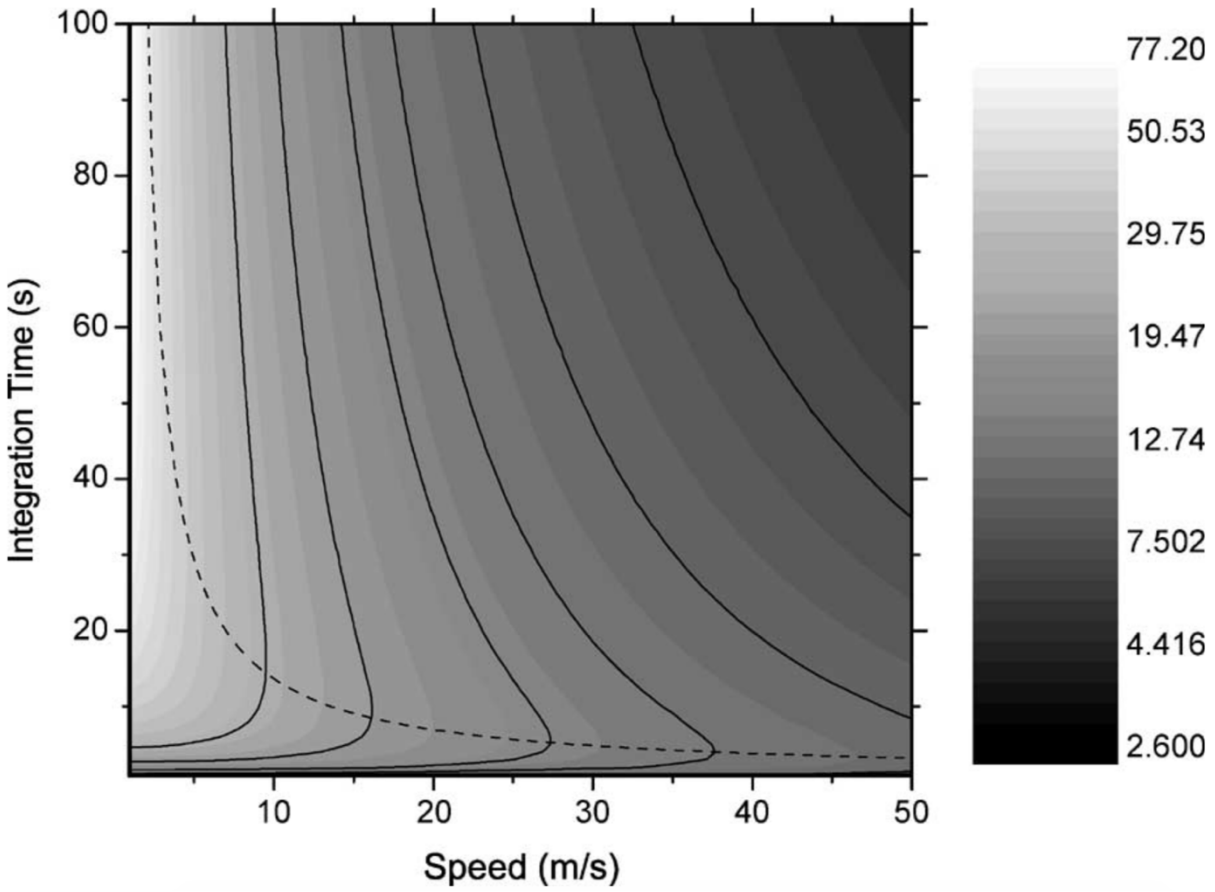


Figure 28: Optimal integration time for arbitrary mobile radioactive source scenario. [39]

\ concern is detecting mobile radioactive materials in vehicles, typical travel paths and distance of closest approach are relatively known as the vehicle is assumed to stay within the lanes of the roadway. Furthermore, if limited information is known prior to deploying a sensor array, the speed limit can be utilized to estimate the optimal timing resolution for the area; 25 miles per hours ($\sim$11 meters per second) for the ORNL sensor array. Using the optimal integration window length relationship described in [29, 96] and assuming a 2 meters and 11 meters per second for the average expected anomaly interactions, an integration time can be acquired; that is, $x = 1.39\, y = 1.39\,(\,2\text{ m}\,) = 2.78\text{ m}$ and subsequently, $t = \frac{2\,(\,2.78\text{ m}\,)}{11\text{ m/sec}} \cong 0.5\text{ sec}$. A relatively short integration time of approximately 0.5 seconds is obtained due to the small distance of closest approach and moderately slow vehicular speeds for the sensor array utilized in this work. In addition, recall that the data in this work are analyzed at a 1 second timing resolution, thus using a sub-second integration time is not addressed here. Using 1 second foreground and 30 second background window widths, six months of data from ORNL sensor array is used to create a ROC curve for the gross count detection algorithm (see page 48 for decision metric). Figure 29 provides resulting ROC curve using the normal approximation error bounds discussed previously. The ROC curve in Figure 29 provides suitable results as an optimal classifier lies in the upper left of the ROC curve with high true positive rate with low positive rate. The normal approximation for uncertainty quantifying decreases at higher decision metrics values (i.e. left side of the ROC curve) as no successful classifications (n_s) are made, resulting in the bounds going to zero. Using a decision threshold at 95% true positive rate, the length of the detected anomalies is provided for context in Figure 30. Approximately 50% of all alarm events detected are equal or less 5 seconds in length. The majority of the longer alarms detected are inherent to monitoring facilities that handle radioactive materials, as normal facility operations, such as

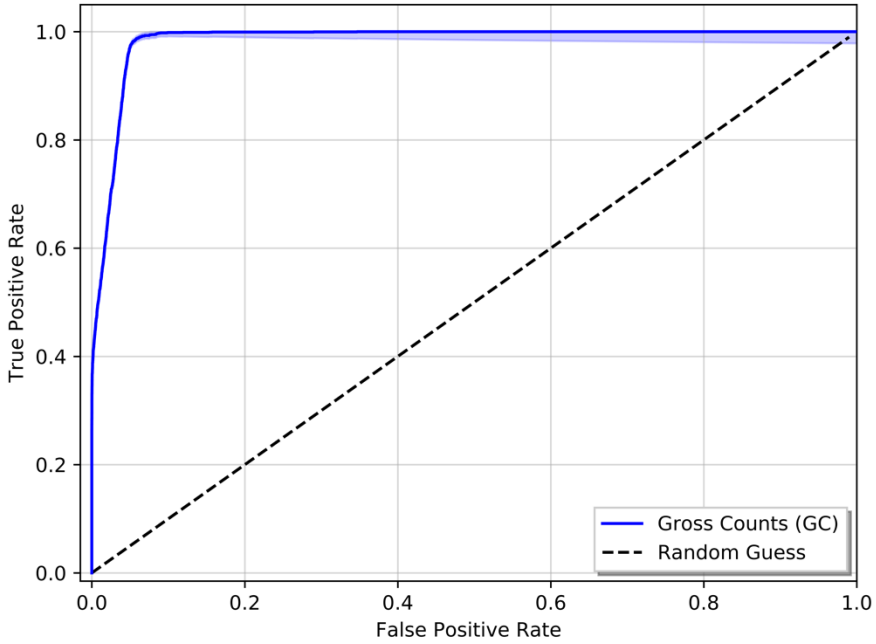


Figure 29: Receiver operator characteristic curve for gross count anomaly detection algorithm.

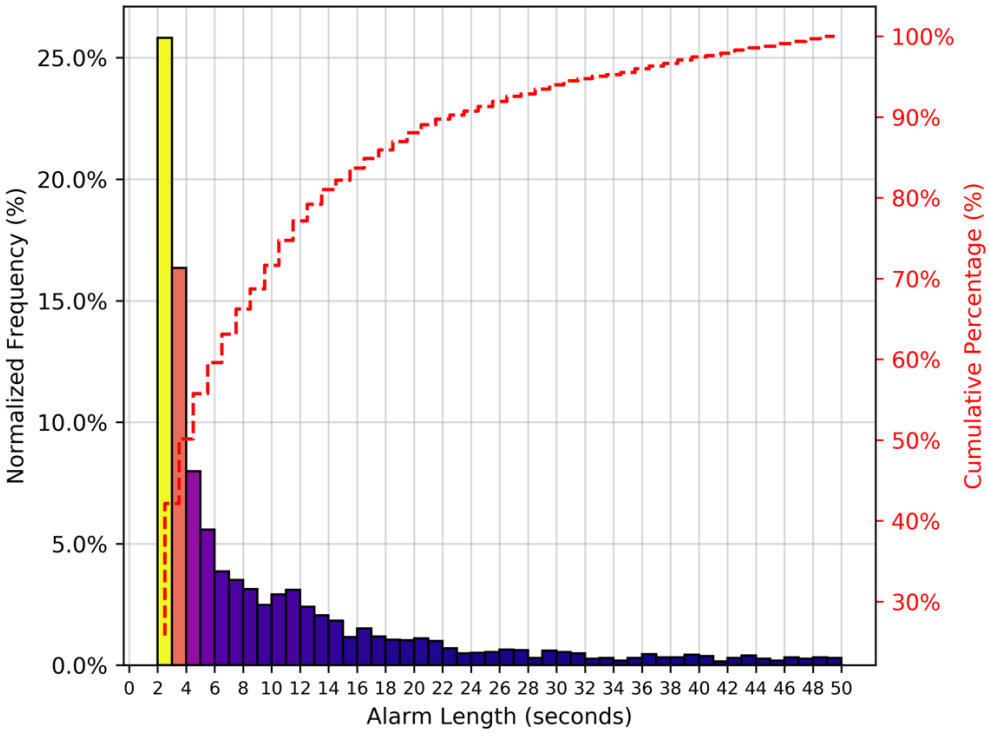


Figure 30: Histogram of alarm length times for anomalies detected from sensor array.

loading or unload materials for transport, require extended periods of time by facility and safety personnel to perform the tasks. An important result to note is that as 50% of the alarms being less than or equal 5 seconds in length provides only a short period of time to collect useful data. For instance, a short interaction can provide sparse energy spectra observed during the alarm events, especially for anomalies with small observed count rates. This will be further assessed in the following chapter with proposed potential solutions.

3.2: LIDAR Data Analysis

To provide contextual information to the radiation anomalies detected using the previously discussed approach, a LIDAR unit can be utilized to detect, track, and, perhaps, classify moving objects in the scene around the nodes. This allows for additional information to be extracted that can be used for approximating properties of the anomaly (e.g., activity of the radioactive material) and other useful inferred features (e.g., velocity estimate).

LIDAR units have been utilized in recent years to provide contextual information for a variety of scenarios, such as aerosol monitoring, geological surveys, autonomous vehicle navigation, and, as shown previously, even radiation detection [97, 98, 99, 100, 101]. In basic terms, LIDAR units operate by counting the time between events, namely backscattered or reflected optical signal, rotating laser. With the photon attenuation of air known prior or approximated using onboard environmental sensors, the distance from the returned signal is estimated from the time delay between emission and receiving the same signal, referred to as the time-of-flight (TOF) principle. The returned signals are typically correlated by emitting a known, specific pattern of pulses by altering the intensity, phase, and/or frequency of the signal [99, 102]. For additional background information on the various methodologies and strategies to perform distance and mapping procedures, the author suggests review the publications [101, 103]

and the text [99, 102, 104] for more in-depth information and comparison of current technologies.

In this work, a Velodyne VLP-16 Puck LIDAR unit that provides data from 16 lasers aligned in a column providing 360° data by rotating the column at frequencies from 1 to 10 Hz [81, 105]. Using the TOF principle, Velodyne provides software to process the data collected in live time and concatenate the data a specific integration time, such as 0.1 second to 5 second, to form 3-dimensional (3D) point clouds. In Figure 31, an example point cloud is provided from a Velodyne LIDAR unit that is shaded from red to blue for low to high confidence returns [106].

It is worth noting that the Velodyne sensors also return an additional parameter referred to as reflectivity that measures the reflective properties of the material or medium the incident beam strikes. This parameter attempts to measure the absorbent and reflectivity properties based on diffuse returns and changes in the returned magnitudes [105]. This parameter is not evaluated in this work and only the 3D locations of the returns are analyzed. The DAQ on the ORNL sensor array store the packets received from the LIDAR units in PCAP (i.e. file extension .pcap) files, in contrast to the more commonly utilized formats, such as LAS (i.e. file extension .las) [107, 108]. The PCAP files contain the timestamp, intensity-of-return, distance, azimuth, and laser unique identification number over a specified time period. To parse the PCAP files, two methods were employed. First, the Robot Operating System (ROS) framework is utilized to read, process, and pipe the down sampled point clouds to Python for further processing (i.e. object detection and tracking) [109]. ROS contains multiple libraries and tools for efficiently reading point cloud file formats, but requires a real-time replay of the saved file; meaning a point cloud file collected over six-hours would have six-hours to read and analyze. Using the ROS libraries provides a pipeline for real-time analysis of experimental or operational scenarios but is not

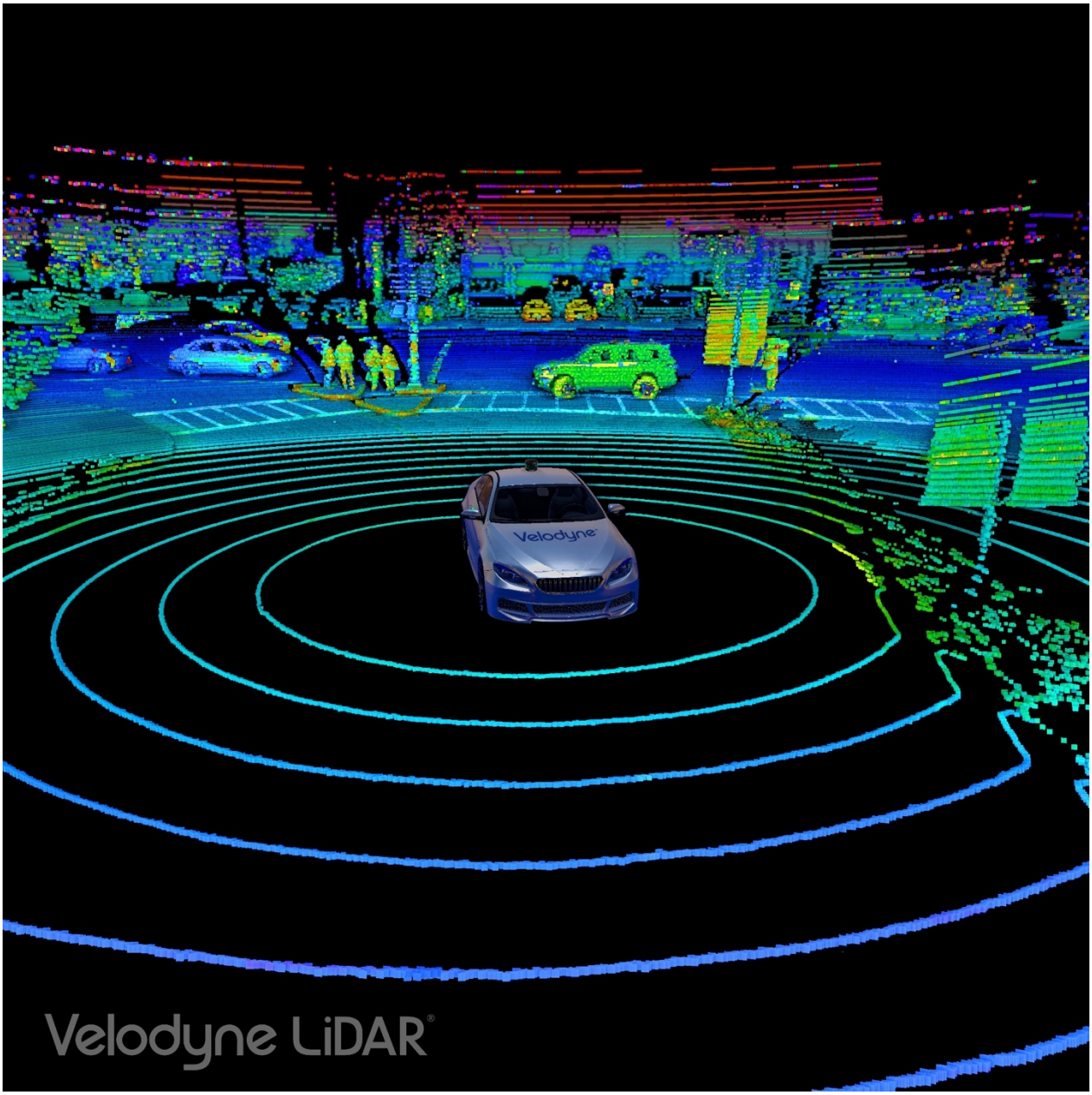


Figure 31: Example 3D point cloud collected from Velodyne LIDAR unit. [106]

utilized for analyzing historic files. To overcome this shortfall, the PCAP files are converted to an HDF5 format which can be manipulated and directly processed in Python. This is less than ideal as it does require a time convert the files but is suffice in this work for analyzing the data in a post-process sense.

Prior to converting the data, the point clouds are down sampled to lower the computation costs in analyzing the large point clouds. It was learned early on during this work that it is important to down sample the data by removing points in specific regions of low important, such as an open field behind a node that observes no traffic. An example point cloud from one of the nodes in the ORNL sensor array is provided in Figure 32 for a 1.0 second integration with a removal of all points behind the Velodyne unit. The LIDAR unit is position in the middle-right of Figure 32 and is assumed to be the origin of the coordinate system. Even with the removal of points in certain regions of the point cloud, the LIDAR units typically observe point clouds in excess of 100,000 points per second when operating at the 10 Hz rotation frequency. In an attempt to lower the computation time required to process the point clouds in near-real time, the point clouds are further processed by performing a ground subtraction along with eliminating low confidence points (i.e. low intensity-of-return). Ground subtraction significantly decreases the amount of point in the points to be processed by removing specific regions in the point cloud, especially dense areas near the LIDAR unit, and is regarded as nonessential for the purposes of detecting and tracking moving objects. Several unique approaches for performing ground subtraction from points clouds can be utilized and references [110, 111] provide more in-depth information.

In this work, a more simplistic approach is utilized by only processing points within a predetermined range along the *z*-axis (i.e. vertical axis with origin at the LIDAR unit). Points

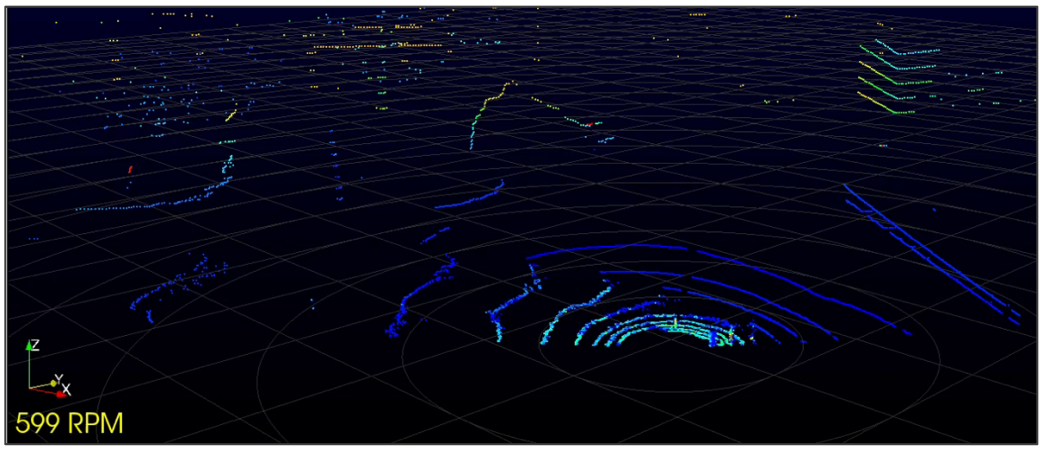


Figure 32: Example image of LIDAR point cloud.

outside the range from 0.3 meters to 4 meters along the *z*-axis are removed along with points greater than 30 meters from the LIDAR unit. The resulting 3D point cloud now only contains relevant region of interests around each node, reducing the point clouds in excess of 100,000 points to less than 30,000 points for 1 second integration. Absent of any robust COTS solution for analyzing 3D point clouds for detecting and tracking objects, the analysis method implemented in this work analyzes the point clouds by creating an occupancy grid by constructing two-dimensional (2D) histogram of the point cloud after collapsing the data along the *z*-axis. Occupancy grids are created by collecting the 2D histogram data for a predetermined integration time and spatial resolution to estimate the probability each cell in the histogram is occupied [112, 113]. Unfortunately, occupancy grids can be computationally expensive to analyze and a trade-off exists between selecting optimal integration time and resolution of the grid (i.e. cell size). A large cell size (e.g., 2 meters by 2 meters) decreases the resolution and subsequently the sensitivity of the detecting a moving object as the points collected could fall within a previously occupied cell. In contrast, a small cell size (e.g., 0.1 meters by 0.1 meters) increases the resolution, but also increases the false-positive rate for detecting a moving object primarily due to large amount of noise inherent to the point clouds. Furthermore, each sensor location has a unique environment in the vicinity of the sensor, thus unique hyperparameters are required for each.

The occupancy grid algorithm alarms when LIDAR points are observed in a previously unoccupied gird cell for three consecutive frames (i.e. three 0.1 second frames). Requiring a previously unoccupied grid cell to have multiple LIDAR points within the cell for three consecutive frames reduced the false alarm rate drastically due to the large noise component inherent to LIDAR point clouds. After a LIDAR detection is made, the LIDAR points in specific

cells-of-interest (i.e. previously unoccupied cell that are now occupied) are clustered and a pseudo-center position of the cluster is calculated by averaging the (x, y) positions of the anomalous points. The center point is subsequently implemented in a Kalman filter to calculate a velocity vector (i.e. expected center position and expected speed in both the x and y directions) until the alarm status returns to normal background via the occupancy grid algorithm [114]. A Kalman filter is particularly suited for this task as it is able to incorporate previous timesteps information to form predictions with associated noise and update, or correct, the predictions upon performing a measurement. The Kalman filter utilized in this manner is extremely common in object tracking applications and is not fully derived in this work, but the author suggests the interested readers to review the references [115, 116, 117]. for additional information for the full derivation along with applications. The output of the filter provides the location and speed of the object, which is utilized to approximate the travel times to the other sensors in the network.

A histogram cell size of 1 square meter (m^2) is utilized in this work to create the occupancy grid as it was found to reduce the point clouds adequately and decreased the false alarm rate to near zero at all six locations. Additionally, even with relatively large points remaining in the point cloud (i.e. ~30,000 at each node), performing the 2D histogram is a relatively low computation expense and can be performed in real-time as only a limited number of cells are after filtering the raw point cloud using the numerical package NumPy in Python [118]. In Figure 33, an overview of the entire process for real-time point cloud processing from importing the raw data stream using ROS and the creation of an occupancy grid in Python after filtering the point cloud. The occupancy grid method is susceptible to errors, especially when performing the background subtraction on non-level terrain (e.g., hillside) and inability classify the detected object (e.g., human, car, or truck).

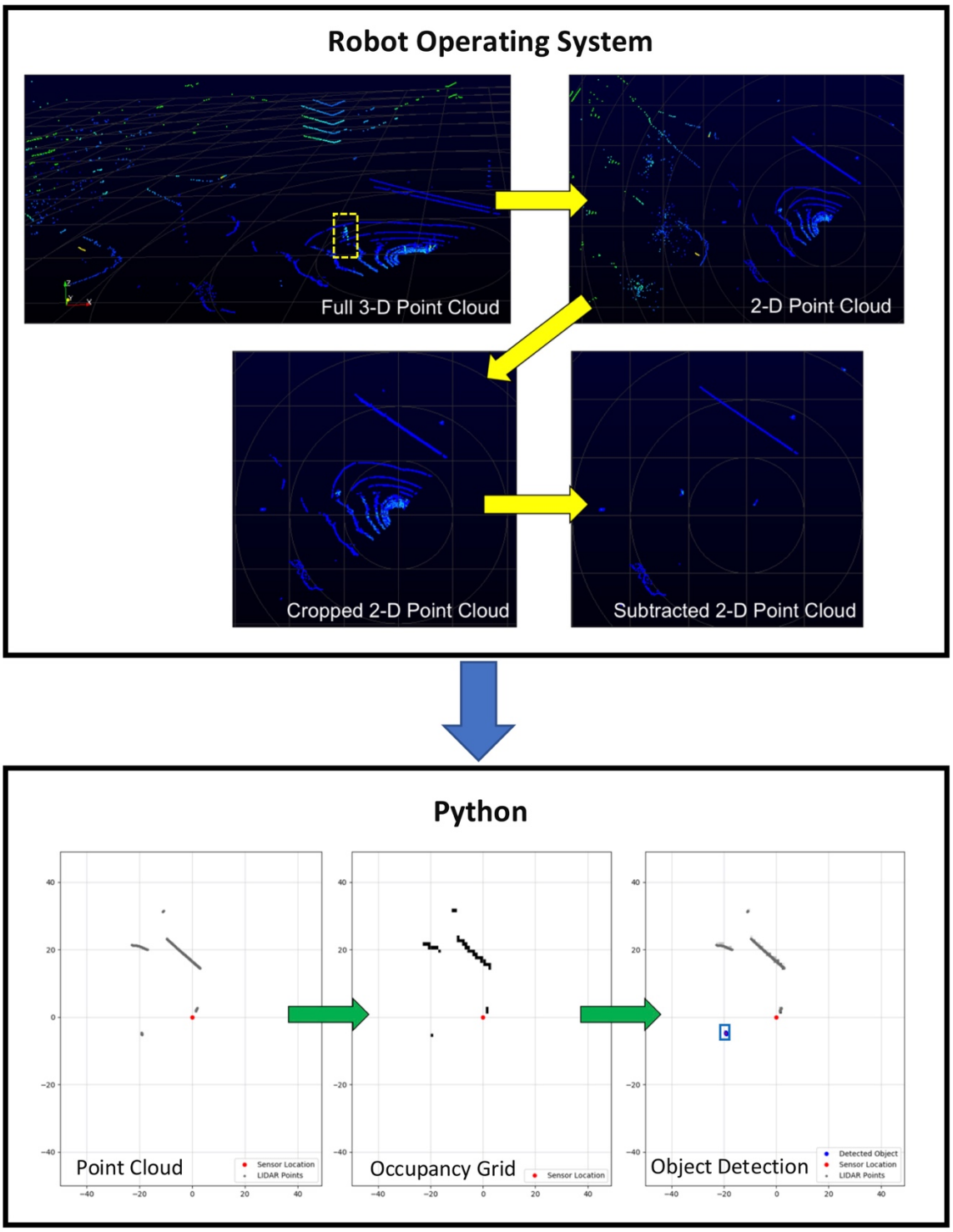


Figure 33: Overview of LIDAR data processing using ROS and python.

3.3: Video Data Analysis

Digital image and video data remain one of the most popular information streams as it is an inherently interdisciplinary with an intersection between optics, sensing, signal processing, and data analytics with plethora of unique motivations within the R&D community [119, 120, 121]. Image and video processing provides a rich data stream that provides information in the immediate environment around the sensor. Three common applications in image and video processing involve object detection, classification, and tracking [121, 122, 123, 124]. These tasks involve transforming an image or video frame into a numerical representation matrix and subsequently iterating over the matrix to locate objects within the image or video frame and classifying said objects. This generalized approach is broadly referred to as computer vision, as the data is typically processed in an automated fashion using trained computational models [125, 126].

Ultimately, the goal of image and video processing for the scenarios in this work is to find objects in the field-of-view of the camera and classify the detected object (e.g., human, vehicle, etc.) for later comparison with other anomalous events. Performing such tasks, while relatively simple for humans, can be nontrivial to automate without human intervention. In an attempt to mimic the process performed in the human brain, neural networks were created consisting of different nodes with multiple layers, similar to the hypothesized architecture of the human brain. The networks can be used to create predictive models in both supervised (i.e. known classification labels) and unsupervised (i.e. unknown classification labels) methods [125, 126]. Neural networks can perform tasks that are invariant to the linearity and parametric relationships in data. The obtained model can provide a robust and generalizable approach for classification, or object recognition. There exists a superfluity of literature on the various techniques, neural

network architectures, and COTS solutions for object detection, recognition, and tracking and the author of this work suggests the reader to review [125, 126, 127] for more in-depth information regarding the fundamental concepts of this problem space.

With the purpose of using COTS algorithms for processing the video data, the chosen approach should be efficient at providing moderately fast results on an edge device with a high degree of accuracy. The variety of unique classes available for classification is not as important for the scenarios described in this work, as the primary objects encountered are expected to be either people or vehicles. With those considerations in mind, there exist several COTS, open source solutions:

1. InceptionV3 [128, 129, 130],
2. RetinaNet [131],
3. Single Shot Detector (SSD) [132, 133],
4. Fast / Faster R-CNN (Region-based Convolutional Neural Network) [134, 135, 136], and
5. You Only Look Once (YOLO) [137, 138].

The speed, or efficiency, at which the specific algorithm operates primarily depends on the resources (i.e. computation power) utilized to process the data and the complexity of the neural network (i.e. number of operations required per image or video frame). There exists a tradeoff that must be understood between a classifiers' accuracy and the processing speed. To quantify the accuracy of an object classifier, the average precision (AP) metric is typically utilized as the de facto standard. The AP combines the precision and recall (see the following equations) of the algorithm's classifications to evaluate the performance of the algorithm in a single metric [17]; $Precision = P = \frac{TP}{TP+FP}$ and $Recall = R = \frac{TP}{TP+FN}$.

In more general language, the precision calculates the accuracy of the classification predictions (i.e. when the algorithm detected an object, how well did it do?) and the recall measures how well the algorithm identifies all, or a portion, of the correct classifications (i.e. how did the algorithm predict every time it should have detected?). The precision is further expanded in regard to object detection using the intersection over union (IoU) metric to evaluate the amount the predicted bounding box overlaps with correct bounding box. This is generalized in the following manner utilizing the area of the predicted bounding box (A_p) and the area of the correct bounding box (A_c) using union (∪) and intersection (∩) set notation with a perfect overlap equal to 1.0, $IoU = \frac{|A_p \cap A_c|}{|A_p \cup A_c|}$ [139]. A minimum IoU value is predetermined prior to training the neural network on a dataset with typical predictions only classified if IoU values are greater than or equal to 0.5 (i.e. commonly reported at AP_{50}) [141]. This means for IoU values greater than or equal to 0.5 are classified as true positives, while values less than 0.5 result in a false positive. Additionally, false negatives are observed when an object is present in the data, but no detection is made by the specific algorithm. The precision and recall are calculated for a variety of confidence thresholds creating a precision-recall curve, which for the readers from the radiation detection community can be thought of as a plot to provide similar information to that of a ROC curve. A large area under the curve for a precision-recall plot provides high recall and precision, where, recall, high recall represents a low false negative rate and a high precision represents a low false positive rate. The AP is calculated by summing the precision-weighted difference between the current and previous recall at the N^{th} threshold value, $AP = \int P(R)\, dR \cong \sum_N P_n\, (R_n - R_{n-1})$.

Since the precision and recall values fall within 0 to 1 range, the AP also falls within the range. The mean average precision (mAP) is generally utilized for reporting an algorithm's

accuracy, where mAP is the averages AP over N number of queries or classifications in the model, $mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i$ [139, 141]. It is important to note that the AP and the mAP can be identical in reporting the value depending on how the authors calculate the metrics; therefore, users should review the methodology implemented for the data and specific model utilized. Fortunately, there exist multiple publications that compare several of the COTS models on identical training and test data for the provided models in the previously shown list [137, 138, 140]. Using the popular COCO dataset of labelled images, the Table 3 borrowed from Redmon, et al. provides a concise comparison of multiple object detection and classification models using mAP_{50} and inference time as performance metric on a Titan X GPU [138]. Of note, the values in the model names in Table 3 correspond to the model size. For example, SSD (321) represents a 321 x 321 input image, which may require resizing of the input image or video frame.

In this work, the YOLOv3 (i.e. Version 3 of YOLO) algorithm is utilized within the Open source Computer Vision (OpenCV) library along with TensorFlow in Python. This model is used due to the model's relatively high mAP and extremely fast inference time (i.e. time required to detect and classify objects) [139, 142]. The YOLOv3 algorithm performs a regression detection model by iterating over the image or video frame for a given bounding box size, sometime referred to as fixed-grid regression [138]. YOLOv3 separates the frame into cells and iterates over the created cells using the trained convolutional neural network and provides classification values. The dimensions of the bounding boxes are predetermined using the dataset to train the anchor box locations within the network rather than naively predicting boxes of all sizes [143, 144]. The YOLOv3 implementation within OpenCV is pre-trained on the COCO dataset and contains 80 unique classes (see page 230 in Appendix for class list) [145]. Unlike the previous versions of YOLO that softmax the class scores and assume classes are mutually exclusive,

YOLOv3 utilizes independent logistic classifiers with the scores above a specific threshold used to classify the specific bound box [138]. An example event from two video feeds on the ORNL sensor array is shown in Figure 34 with the resulting bounding box for the detected object.

No performance issues were detected when implementing the YOLOv3 model regarding the inference time and object detection on the ORNL sensor data, but an observation was made when attempting to select the object classification when processing time series data for an event (e.g., motion event). During an event, the model processes each frame in the video feed and provides an output vector of confidence for each unique classification node in the output layer. Conventional methods select the highest confidence classification in the output layer as the predicted class for the detected object in scene, but during a motion, different high confidence classifications can occur that contradict one another. For example, it was not uncommon to observe a misclassification of an object at the edges of the frames (i.e. when entering or leaving the camera field-of-view) or if the view of the object become partially obstructed. If the conventional approach is employed, the classifications can vary throughout event, as shown in Figure 35, making it difficult to select a classification. In Figure 35, the object enters the scene and the following classification vector for the 3-frame motion event is obtained, $V = [car, truck, truck]$.

To select the correct object classification for the motion event, two simple methods can be utilized. First, if the user trusts the model's output, the classification with the highest confidence can be used to classify the object for the event (e.g., truck with a value of 0.89 in Frame 3 in Figure 35). Second, one can use the mode of the classification vector as it appears most often. For the example in Figure 35, both methods would provide the same result, but is not the case for all events. In this work, the mode of the output classification vector is utilized as this

Table 3: Performance comparison of various object detection and classification models. [138]

Model	$\boldsymbol{mAP_{50}}$	**Inference Time (msec)**
SSD (321)	45.4	61
SSD (513)	50.4	125
Fast R-CNN	59.1*	172
Faster R-CNN	51.9	85
RetinaNet-50 (500)	50.9	73
RetinaNet-101 (500)	53.1	90
RetinaNet-101 (800)	57.5	198
YOLOv3 (320)	51.5	22*
YOLOv3 (416)	55.3	29
YOLOv3 (608)	57.9	51

* Optimal value for specific parameter.

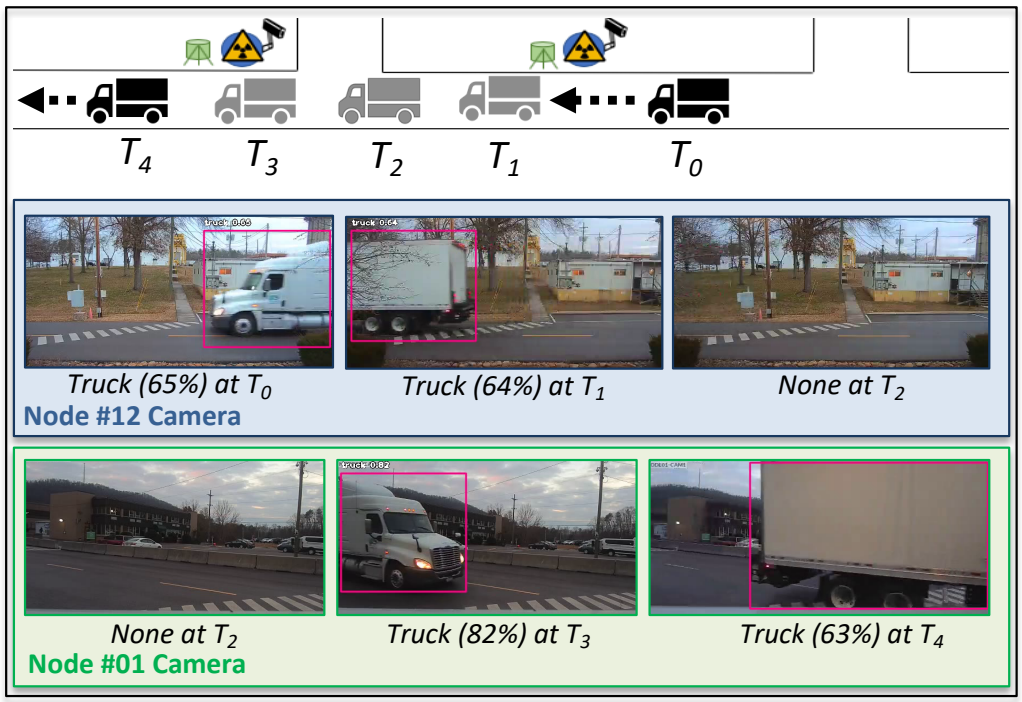


Figure 34: Time series video processing from sensor ORNL using YOLOv3.

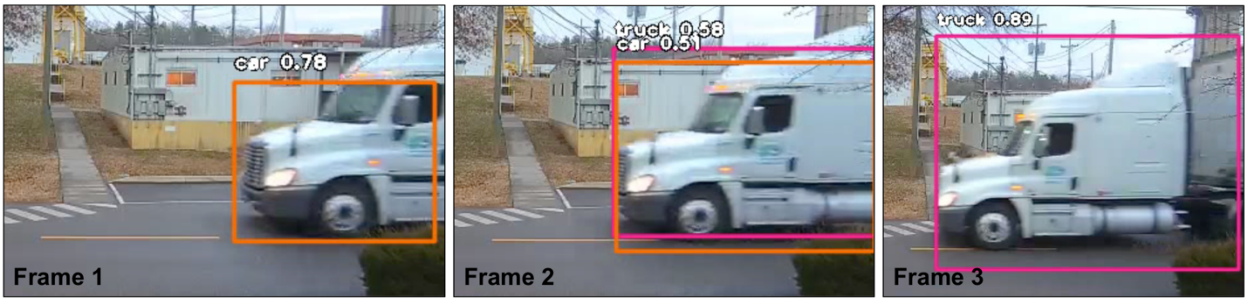


Figure 35: Example model output from motion event with differing classifications.

method assumes that the correct classification will be most dominant in the output classification vector. It was found during experimentation that this claim was true for the cameras on the ORNL sensor array, primarily due to the location and field of view of the cameras with respect to the walking paths and roadways. An alternative approach was utilized that averaged the classification confidences for each reported, unique classification and subsequently selecting the classification with the highest mean confidence. This technique provided identical results to the mode selection technique during experimentation, but will be utilized moving forward as it thought to provide a better representation of the actual object classification from the event.

The YOLOv3 model is further customized by only returning only the unique classes of interest, which in this work is human, car, or truck classifications (i.e. all remaining classifications are set to zero). Of note, this modification does not provide any speed improvement when performing the detection and classification calculation but limits the classifications to only specific classifications expected to be transporting radiological materials. Furthermore, a smaller neural network version of the YOLOv3 architecture is used in order to decrease the inference time to detect and classify objects by decreasing the depth of the convolution layers from 53 to 19 (i.e. DarkNet-53 to DarkNet-19 architecture). [138] The tiny-YOLOv3 does lower the classification accuracy of the model when compared to the architecture in YOLOv3, but the real-time performance and less strain on the system from a computation standpoint is deemed more vital.

4: Distributed Radiation Detection

To quantify the benefit contextual sensors provide to tracking anomalies through a sensor network, it is first important to first understand the ability and limitations of a radiation-only sensor network. In the absence of any available literature on truly distributed radiation sensor networks, this work presents two approaches that can be utilized in unison for tracking anomalies. The first utilizes spectral information collected during the radiation anomaly and compares the collected spectrum with subsequent radiation anomalous spectra collected at down-stream sensors. This approach provides quantifiable evidence that the anomalous spectra collected at multiple sensors originated from the same radioactive source distribution. The second approach presented utilizes the spectral information to create a region-of-interest approach to improve down-stream detection sensitivity. The region-of-interest is selected by obtaining the highest signal-to-noise ratio region from the initial anomalous energy spectrum. This approach expands upon a reported censor energy window method to provide a more robust approach [146, 147].

In this chapter, both approaches are first presented along with analysis on multiple methodologies to accomplish their respective task. The down-selected methods for both approaches are subsequently employed on synthetic and anomalous radiation event data from ORNL sensor array to evaluate the methods' output. The final section of this chapter attempts to tie the former sections together by providing a more in-depth assessment by providing a decision-making architecture using the previous output on the of the ORNL sensor array data to provide an end-user with a detection decision in an automated fashion.

4.1: Energy Spectra Comparison

Upon detecting an anomalous signal from the radiation detector, the spectral information during the alarm event can be utilized for several reasons. If the source intensity is relatively high compared to background and high energy resolution detection system is employed, then radioisotope identification methods can be employed. Conventional radioisotope identification methods utilize precalculated energy spectra templates for various radioisotopes and compare with the collected energy spectrum with the template [148, 149]. In an operational setting, radioisotope identification can be difficult due to short interaction times, weak source activity, or lack of signal in the photopeak due to the presence of shielding or attenuating material between the source and detector. Conventionally, radiation detection systems with low to moderate energy resolution (e.g., NaI(Tl) or PVT) detectors are utilized, further increasing the difficulty of the radioisotope identification due to the broad energy peaks. Worth noting, spectra from low resolution detectors have been able to accurately identify source in low signal-to-noise ratio operations by deconvolving the spectral data. Performing the deconvolution filtering process requires full knowledge of the detector's response, which is typically acquired by modelling the response using particle transport codes (e.g., Monte Carlo N-Particle code, or MCNP) [150]. In this work, the NaI(Tl) radiation detectors utilized are low energy resolution and primarily observe mobile radioactive source moving at high rates of speed, thus no radioisotope identification process is employed.

If partial or imperfect spectral information is obtained and radioisotope identification is difficult to accurately obtain, the collected anomalous energy spectra can be compared to spectra collected at other detection systems. This process assists in tracking a source in an area by measuring the similarity between the collected anomalous signals. For instance, assume a

relatively strong source is detected at a location and no radioisotope identification is made. The energy spectra collected during the anomalous event can be combined and background-subtracted to obtain an, often very noisy, approximation of the source spectra. If the same source subsequently passes another detection system, the apparent source spectra measured during the previous anomalous event can be compared to the current background-subtracted spectra to verify if this is the same radioactive material or a new material all together. To provide additional context to this example, a source spectrum was modelling in MCNP with gaussian broadening employed for two different source activities combined with a normalized background spectrum obtained from the ORNL array; both distributions are shown in Figure 36.

Assuming the radioactive source is initially detected at Node01 and, after some amount of time passes, is detected at Node02, the spectra during the events can be summed, as shown in Figure 37. Both spectra obtained from the MCNP models are sampled as Poisson variables to align with realistic samples. Moving forward, if the radiation background is relatively stable for both Node01 and Node02 systems, a background spectrum can be subtracted from the anomalous spectra which should be analogous to a sample distribution from the source spectrum, show in Figure 38. Two issues arise when attempting to compare the resulting background-subtracted spectrum. First, if the anomalous events typically occur quickly and the radioactive material has a low to moderate activity, the resulting spectrum collected can be sparse and noisy. Second, the intensity observed can be can vastly differ, primarily governed by the inverse-square law, timing resolution, and speed of the source, which creates unique issues when comparing the spectra using distance-based measures

To overcome the sparse and noisy nature of collected spectrum, two approaches can be utilized:

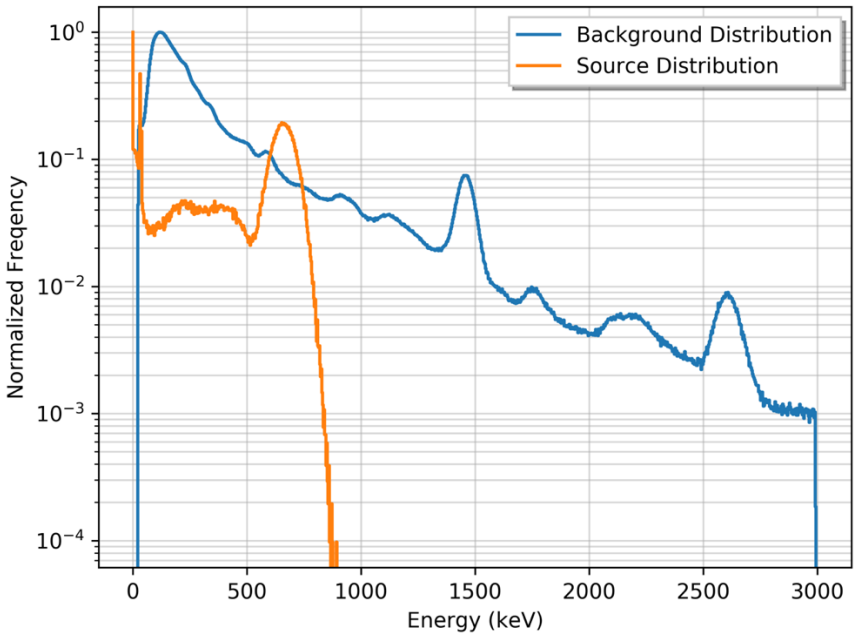


Figure 36: Example distributions for background and a radioactive source.

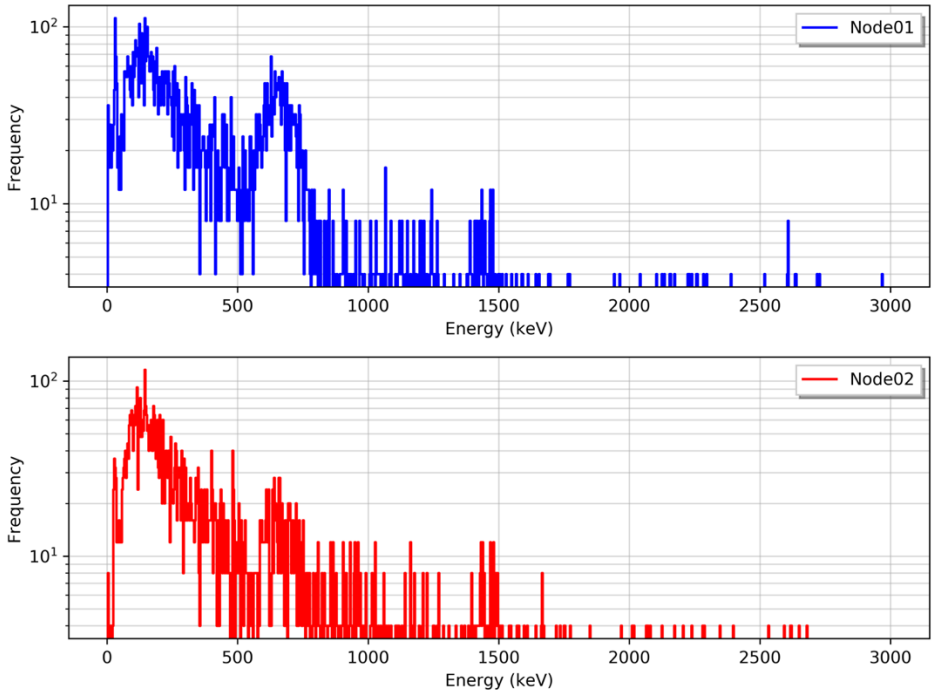


Figure 37: Source spectrum from two locations. (Top) Example energy spectrum collected during anomalous event at Node01. (Bottom) Example energy spectrum collected during anomalous event at Node02.

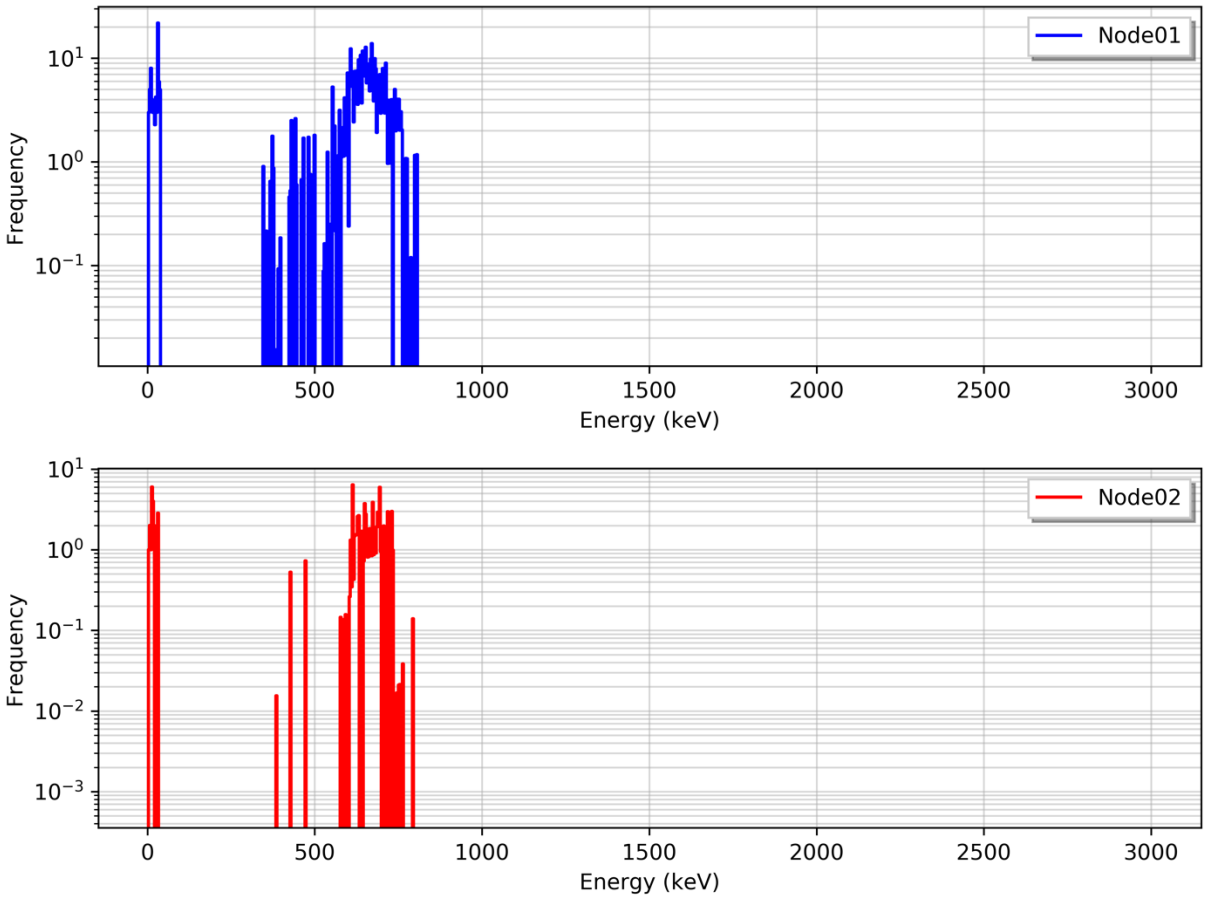


Figure 38: Background subtracted energy spectra for Node01 (top) and Node02 (bottom).

1. modify the spectrum binning structure, or
2. apply a filter to smooth the spectra.

Applying a new binning structure to the spectra has benefits to lower the signal-to-noise ratio in specific region where signal from the source is much greater than the background component, but resolution is sacrificed. A common solution is to linearly combine, or summate, the bins in the spectrum to a lower order (e.g., 1024 bins to 128 bins). The second approach applies a filter to decrease the noise while preserving the overall shape of the spectra by smoothing the data. With a plethora of filtering algorithms available, there exist multiple methods to smooth noisy data, such as polynomial smoothing splines, moving average convolutions, or kernel density functions. It is key that the filter chosen should conserve the underlying statistics, where the area under the curve remains equal with no values are lost and nonphysical values are possible, such as values at negative energies. If a moving average, or windowed, approach is utilized, the size of the window as well as the regression method used are key hyperparameters to predetermine to match the user case. In the case of radiation spectra, these hyperparameters should account for the underlying energy resolution by selecting a larger window for low to moderate energy resolution systems (e.g., NaI(Tl) or CsI(Tl)) and a smaller window for high energy resolution systems (e.g., LaBr(Ce), CZT, HPGe). Furthermore, care should be taken when selecting the polynomial order utilized during regression or spline methods, where nonphysical values can occur in sparse regions of the spectrum.

In this work, a popular digital filter method known as the Savitzky–Golay filter is utilized to smooth the spectra. This filter convolves the spectra by applying a low-degree polynomial to successive subsets of the spectra via linear least squares regression. A moving average method is applied to the spectra to smooth the high frequency bins to populate the counts starved, sparse

regions of the spectra. To optimize the window length and polynomial order hyperparameters, the windows size is varied from 5 to 101 for a first and second order polynomial model. The window size is required to be an odd integer as an even integer length window forces an undesired shift in the filtered signal, since the filter is applied on a delay of $(N-1)/2$ for N samples. This is similar to the reasoning for utilizing odd-length windows for a moving-average filter. Using these parameters, the filter was applied to the background-subtracted spectra at Node01 and Node02 in the previous example.

It has been shown in other works that the estimated variance decreases as the window length increases, but also increases the bias error at the same time [150]. This relationship remains true until a particular window length threshold is met where the variance begins as the length is too large; therefore, there is a tradeoff when selecting the design hyperparameters. For this analysis, a quadratic model is chosen as it provides a higher degree of smoothing in the sparse regions of the spectra for both nodes and a window length of 51 is selected when using the original 1000 bin structure. Using these design hyperparameters, Figure 41 provides the resulting filtered spectra for both nodes.

It has been shown in other works that the estimated variance decreases as the window length increases, but also increases the bias error at the same time [150]. This relationship remains true until a particular window length threshold is met where the variance begins as the length is too large; therefore, there is a tradeoff when selecting the design hyperparameters. For this analysis, a quadratic model is chosen as it provides a higher degree of smoothing in the sparse regions of the spectra for both nodes and a window length of 51 is selected when using the original 1000 bin structure. Using these design hyperparameters, Figure 41 provides the resulting filtered spectra for both nodes.

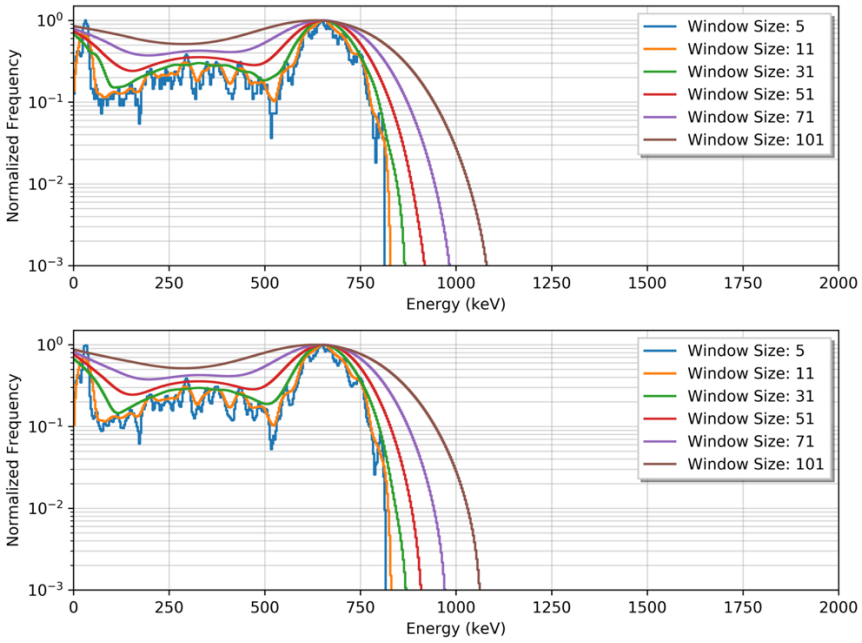


Figure 39: Filtered spectra for Node01. (Top) Filters of various window lengths using a linear model. (Bottom) Filters of various window length using quadratic model.

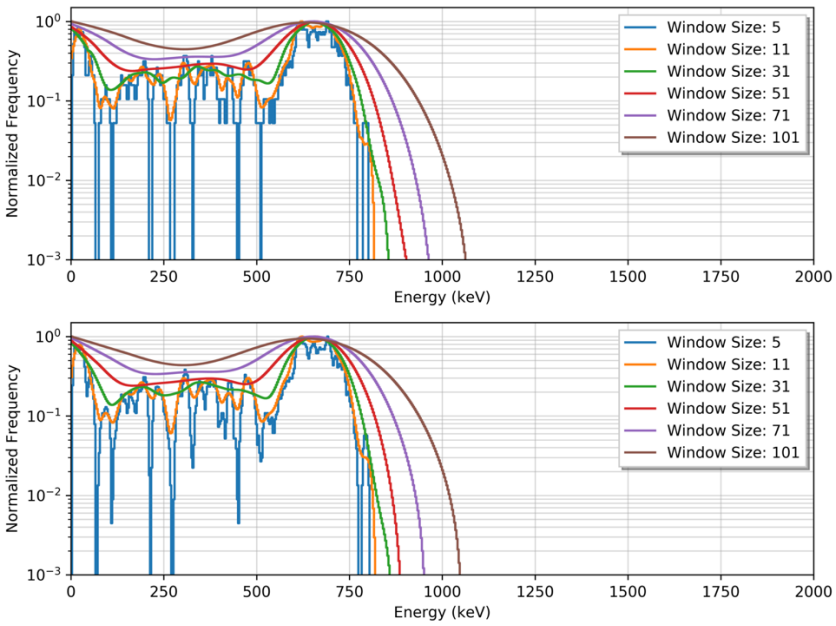


Figure 40: Filtered spectra for Node02. (Top) Filters of various window lengths using a linear model. (Bottom) Filters of various window length using quadratic model.

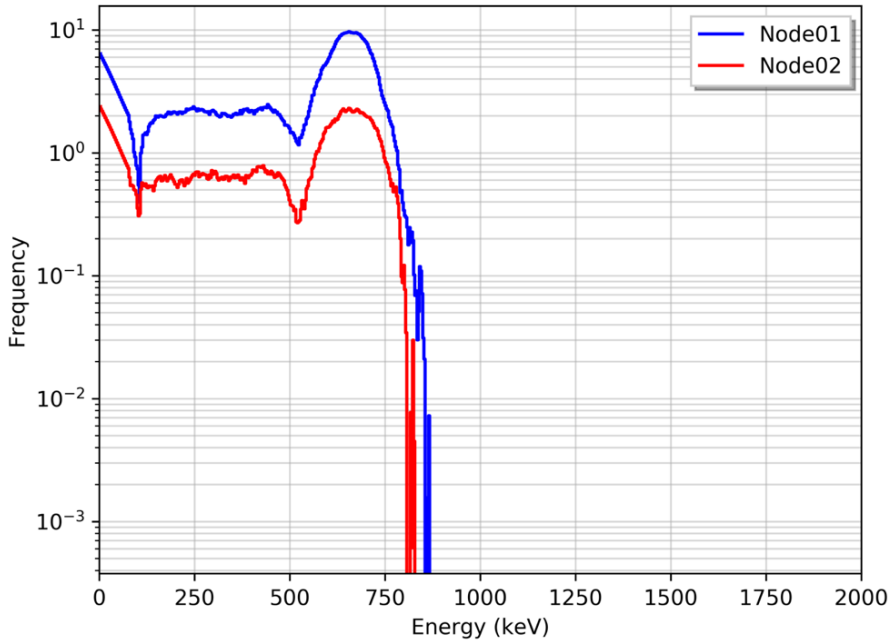


Figure 41: Background-subtracted and filtered spectra for both example anomalous spectra.

Now that sparsity of the spectra has been decreased by smoothing each spectrum by applying the Savitzky–Golay filter, the second issue previously highlighted regarding comparing the spectra was the difference in intensity between spectra. This can create an issue when comparing the spectra if a distance metric (e.g., L^1 or L^2 norm) is used to provide a quantitative comparison between the two spectra. To overcome this issue, one can either normalize the spectra to the same scale to provide relatively equal frequency values yet maintain the overall shape or use a comparison method that is scale invariant. Two such approaches to normalize or standardize spectrum is the divide-by-maximum and the sum-to-unity methods. The divide-by-maximum method normalizes the spectrum by dividing each value by the maximum value of the spectrum, while the sum-to-unity approach normalizes the spectrum by dividing each value by the summation of spectrum. Both approaches were applied to Node01 and Node02 filtered spectra, as shown in Figure 42.

With the spectra on similar scales, multiple distance-based models can now be used to quantify the similarity between spectra with results provided later in this report. It is important to note that there exist two key properties that are vital when comparing energy spectra. First, the detector systems should be relatively identical, or the vector comparison can fail as physical properties that contribute to each energy spectrum will differ (e.g., energy resolution or intrinsic efficiency). Secondly, the spectra should have near-identical binning structure. In practice, the calibration can change due to various operational parameters. Methods exist for performing online stabilization, which can assist in verifying that each element in the spectra being compared correspond to the same energy range [151, 152]. Furthermore, provided a known energy calibration, the spectra can be binned to a user-specified binning structure to verify near-identical binning structures for the compared spectra.

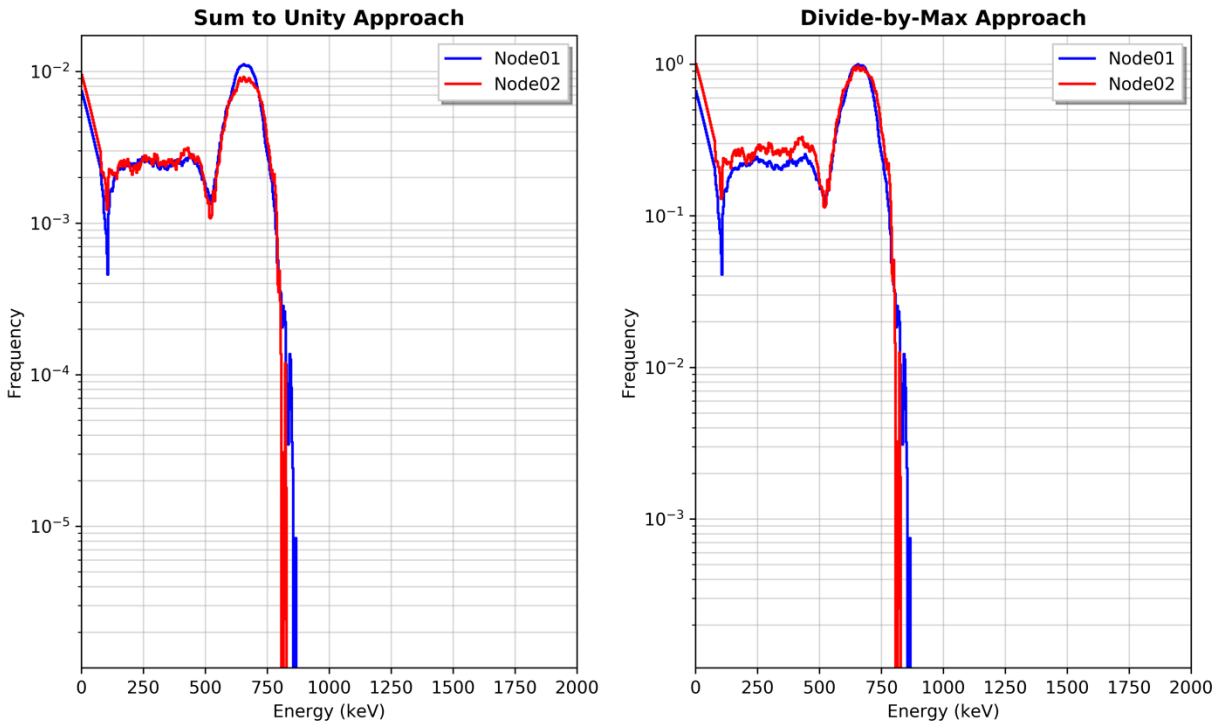


Figure 42: Normalized and filtered spectra using divide-by-max and sum-to-unity approach.

There exist several models that quantifies the similarity between two vectors. In this work, several comparison methods are evaluated: chi-squared test, vector norms (i.e. L^1 and L^2 norm), Kolmogorov–Smirnov test, and cosine similarity. Each of the models are briefly introduced and later applied to the example anomalous spectra to compare each model using both the initial filtered spectra and the normalized, filtered spectra.

4.1.1: Chi-Squared Test Methodology

A chi-squared test is possibly the most common and well-known method for performing a goodness-of-fit test and testing relationships between parameters. This is accomplished by performing a hypothesis test and calculates a chi-squared (χ^2) test statistic value to evaluate the hypotheses. The conventional chi-squared test of independence assesses if a significant relationship between variables exists, which can be utilized in this case if one sample is assumed as the expected frequency distribution. The test statistic is then calculated and compared to an expected χ^2-distribution for a given number of degrees of freedom, most commonly calculated as one less than length of vector. The conventional χ^2 statistic commonly used is calculated by taking the squared differences between the observed and expected data divided by the expected frequency for that bin and summing over all bins, $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$. Given a specific significance level (α), the calculated test statistic can be compared to the critical value in a chi-square table. If the test statistic is greater than the critical value at the specific significance level, than there exists a significant difference between the observed and expected vectors. A more conventional notion is that large χ^2 values indicate the null hypothesis (i.e. the observed data is drawn from the expected distribution) is relatively unlikely.

In this work however, a two-sample test approach is employed as it provides a method. The two sample test states that the two samples are drawn from the same distribution, while the

alternate hypothesis states the samples are not drawn from the same distribution. Note that this test does not explicitly state the form of the underlying distribution, only tests if the samples are drawn from the same distribution. For the chi-squared two sample test, the test statistic is calculated as follows, $\chi^2 = \sum_{i=1}^{k} \frac{(K_1 R_i - K_2 S_i)^2}{R_i + S_i}$ where, the summation is performed over the two k-length samples, R and S and K_1 and K_2 are scaling constants when the two samples have unequal number of data points. K_1 and K_2 are calculated as follows $K_1 = \sqrt{\frac{\sum_{i=1}^{k} S_i}{\sum_{i=1}^{k} R_i}}$ and $K_2 = \sqrt{\frac{\sum_{i=1}^{k} R_i}{\sum_{i=1}^{k} S_i}}$, respectively. In cases presented in this work, two energy spectra are being compared with identical binning structures. Therefore, the scaling constants are no longer required and can be removed, resulting in the following two-sample chi-squared equation, $\chi^2 = \sum_{i=1}^{k} \frac{(R_i - S_i)^2}{R_i + S_i}$.

Upon obtaining a χ^2 value, a chi-squared percent point function for given significance level (α) and degrees of freedom is calculated numerically, similar to the conventional chi-squared test. In this case, the degrees of freedom value is obtained by subtracting the number of filled bins (i.e. bins with nonzero values) by one. The reason being is that if both samples contain no value in a bin, a divide-by-zero error is introduced in the previous equation. Using this method, the chi-squared value supplies a *p*-value which, given a set significance value, the test can either accept or reject the null hypothesis, which is how the decision when comparing two spectra will be made.

4.1.2: Vector Norm Methodology

A norm is a mathematical function that with specific properties of scalability and additivity, which are not addressed in this work. For more information on norms, the author recommends a review of [154]. A norm function, $|x|_p$, referred to as *p*-norm, is used in a variety of machine learning operations to measure the distance between vectors. The *p*-norm function of

a single vector is provided in the following equation, where parameter p represents the order of the norm, such L^1 and L^2 norm $|x|_p = (\sum_i |x_i|^p)^p$. The L^1 norm, where p is equal to 1 in this case, is a natural way of measuring the distance of between vectors as it calculates the sum of absolute difference of the components of the vectors. It remains a useful method for many applications, such as image processing and compressed sensing, as it is robust against extreme or outlier values. as represented in the following equation for two n-length vectors $X = [x_1, x_2, \dots, x_n]$, $Y = [y_1, y_2, \dots, y_n]$, and, subsequently, $L^1 = |X - Y|_1 = (\sum_i |x_i - y_i|^1)^{\frac{1}{1}} = |x_1 - y_1| + |x_2 - y_2| + \cdots + |x_n - y_n|$. This norm function is informally referred to as the Manhattan Distance or Taxicab norm as the distance between two points resembles the path a vehicle, or taxi, would drive in an urban area, such as Manhattan. Another way of stating this is that this method is the sum of the lengths of a line segment between two points, as depicted in Figure 43.

The L^2 norm measures the distance by calculating the square root sum of the squares of two vectors. This norm function is commonly referred to as the Euclidean norm or Euclidean distance, representing the shortest distance between two points and is invariant under rotation Using the same vectors as the L^1 norm equations, the following equation provides the L^2 norm, $L^2 = |X - Y|_2 = (\sum_i |x_i - y_i|^2)^{\frac{1}{2}} = \sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2 + \cdots + |x_n - y_n|^2}$. Due to each component in the vectors being squared, outliers are given additional weight, potentially skewing the results, compared to L^1 which measures outliers linearly. In a more general sense, the larger the p-value in the norm function, the more important and weight the largest single feature distance becomes.

Recall that a primary issue when using a distance-based approach for comparing the vectors is to obtain an optimal method for standardizing or normalizing the vectors. After the

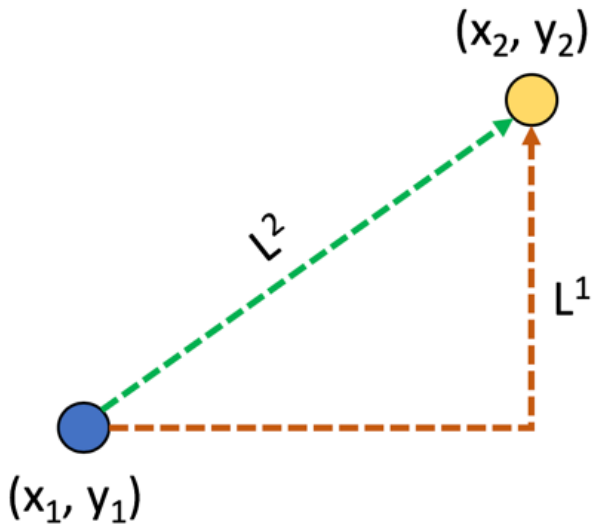


Figure 43: Illustration of L^1 and L^2 norm distance measures.

distances are calculated using a specific method, such as the L^1 or L^2 norm, the user must quantify how the distance method output correlates to vector similarity. This is data-specific as well as situational-specific with potential for increased error or misclassifications, such as poor vector standardization. For this reason, more robust methods should be utilized.

4.1.3: Kolmogorov–Smirnov Test Methodology

The Kolmogorov-Smirnov test, often referred to as the K-S test, is a nonparametric test meaning that this method does not require the data to follow a Gaussian, or normal, distribution, unlike the parametric tests such as the *t*-test, ANOVA, and *f*-test [155]. This is a key property as no assumptions regarding the underlying distributions for both samples are made. The K-S test is utilized to compare the overall shape of two sample distributions by performing a hypothesis test. The null hypothesis states that the two samples, or spectra in this case, are drawn from the same distribution. The K-S test performs a hypothesis test by first calculating the cumulative density function (CDF) of the two sample distributions and measures the difference, element-by-element between the two CDFs. In this instance, the sum-to-unity, filtered spectra are used as the probability density functions to create the CDF, as shown in Figure 44. The K-S test uses test statistic, *D*, to quantify the similarity between the two CDFs. This parameter is equal to the maximum absolute difference between the two CDFs. As such, the K-S test remains robust and relatively sensitive to both the amplitude and the shape of the distributions of the CDFs. If the two distributions were exactly identical, the two CDFs would lie on top of one another and the *D* test statistic would be zero, near impossible in practice. Ideally, an extremely small test statistic would indicate that the two samples are drawn from the same distribution.

This test statistic alone does not provide enough information to perform the hypothesis test previously presented. To accomplish this, a K-S Table is utilized to calculate the *p*-value

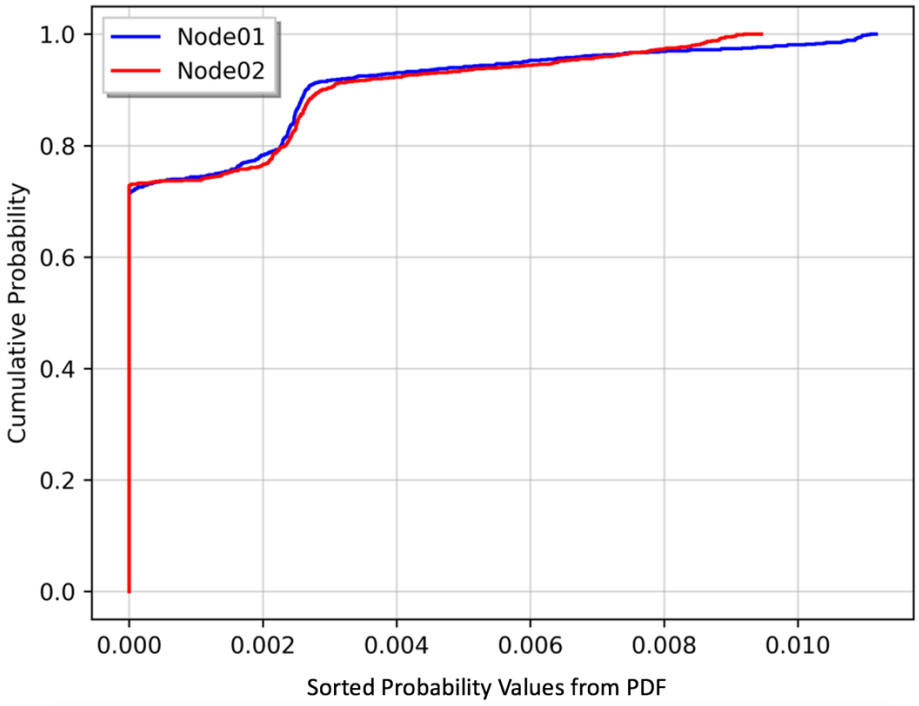


Figure 44: Cumulative density functions for Node01 and Node02 anomalous spectra.

representing the probability of observing the test statistic, D, at least as large as the calculated test statistic. Kolmogorov populated such a table for the expected distribution of the test statistic if the null hypothesis (i.e. both samples are drawn from the same distribution) is true. A more general approach to hypothesis testing rejects the null hypothesis if the following equation is true given a predetermined rejection level, α (i.e. Type I error), $D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{n*m}}$. In this equation, n and m are lengths of the two vectors, $D_{n,m}$ is the calculated test statistic, and $c(\alpha)$ is follows the relationship, $c(\alpha) \cong \sqrt{-\frac{1}{2}\ln\left(\frac{\alpha}{2}\right)}$. For large length vectors, these two equations can be utilized in the hypothesis testing without using the p-value tables created by Kolmogorov. This approach along with table lookup p-value will be used when evaluating each vector comparison approach.

4.1.4: Cosine Similarity Methodology

The cosine similarity approach has gained popularity in recent years as machine learning and data analytics have become more heavily utilized in various industries. This approach differs from the approaches previously discussed, as no standardization or normalization of the two vectors being compared is required; a fundamental flaw in the other approaches. That is, the cosine similarity measures the angle between two vectors in a multi-dimensional ($\mathbb{R}^N$) space, irrespective of their magnitudes, where a small angle represents high similarity. Figure 45 provides an illustration of this concept in a three-dimensional ($\mathbb{R}^3$) space. From this illustration, the angle between the two vectors, denoted by the red curve, represents the value measured by the cosine similarity. The cosine of two vectors is calculated by dividing the Euclidean dot, or inner, product of the vectors by their norms (i.e. commonly referred to as magnitudes or lengths) as expressed as $\vec{a} \cdot \vec{b} = \|\vec{a}\| \left\|\vec{b}\right\| \cos\theta$ and $\cos\theta = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\| \|\vec{b}\|}$. Where, the inner product is the sum

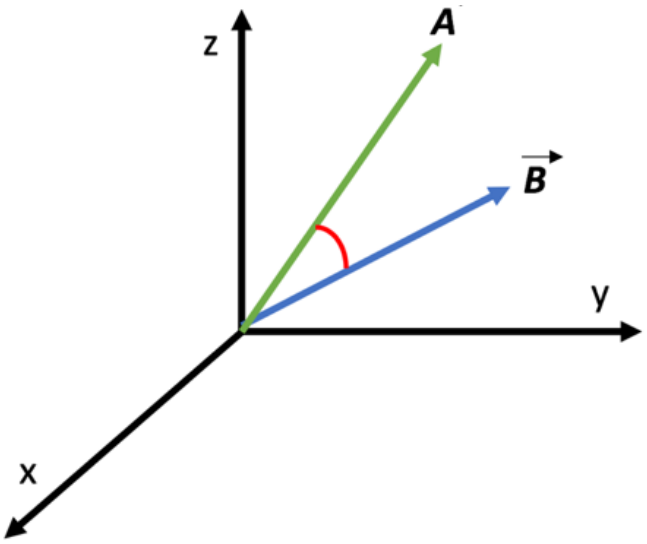


Figure 45: Illustration of cosine similarity concept comparing two vectors.

of the products of the components for each vector and the norms is the square root of the dot product of the vector, expanding the parameters $\vec{a} \cdot \vec{b} = \sum_1^n a_i \cdot b_i$, $\|\vec{a}\| = \sqrt{\vec{a} \cdot \vec{a}} = \sqrt{\sum_1^n a_i^2}$, and $\left\|\vec{b}\right\| = \sqrt{\vec{b} \cdot \vec{b}} = \sqrt{\sum_1^n b_i^2}$. Combining these equations back into the original cosine equation, the following is obtained and provides the relationship that will be used in this work, $\cos\theta = \frac{\sum_1^n a_i \cdot b_i}{\sqrt{\sum_1^n a_i^2}\sqrt{\sum_1^n b_i^2}}$. This final equation is used to calculate the cosine similarity between two vectors. The resulting values range from [-1.0, 1.0] representing exactly opposite and exactly identical, respectively, with a value 0.0 indicating orthogonality. Thus, in this work, if the background-subtracted, anomalous spectra from multiple nodes have a cosine similarity near 1.0, then it is expected that the same radioactive material triggered each. Furthermore, negative values returned from the cosine similarity method is physically impossible in this case as that would require negative values in the spectrum, thus the expected value range is reduced to [0.0, 1.0]

4.1.5: Preliminary Analysis of Comparison Methodologies

Each of the vector comparison methods discussed previously is utilized on the example anomalous spectra provided in the beginning of this section. To provide a more comprehensive analysis, each approach is evaluated on three spectra from both nodes (i.e. Node01 and Node02):

Spectra Type 1: background-subtracted spectra,

Spectra Type 2: background-subtracted, filtered spectra, and

Spectra Type 3: background-subtracted, filtered, sum-to-unity normalized spectra.

The results for the vector norm methods (i.e. L^1 and L^2 norms) are presented in a separately to limit the reader from directly comparing the results with either the two-sample chi-squared

test, K-S Test, or the cosine similarity methods. As previously stated, the vector norm methods employed in this work provide a distance measure between two vectors and a distance threshold for accepting or rejecting the comparison must be independently acquired for the specific user-case. All the methods improve as the spectra are smoothed and standardized by applying the Savitzky-Golay filter and sum-to-unity normalization. This is shown by a decrease in distance for the norm-based methods and decrease in the test statistic for both the two-sample chi-squared test and K-S Test. The cosine similarity observes an increase after the filter is applied to the background-subtracted spectra but shows no improvement after normalization. Recall that the cosine similarity is invariant to the vector magnitudes, therefore no normalization is required as shown in this example.

A primary disadvantage of the distance based L^1 and L^2 norm approaches is selecting a threshold for the specific test statistic. For this reason, a more robust, generalizable solution for comparing vectors will be chosen, such as the methods in Table 5. Of the remaining methods, the cosine similarity provides the more robust solution, as no normalization is required, and the calculated value provides an intuitive way to quantify the similarity between two spectra (i.e. recall, a value of 1.0 represents similar vectors). For these reasons, the cosine similarity method will be utilized for comparing anomalous spectra when a radioactive source moves through the distributed sensor network.

4.2: Region-of-Interest Detection

Provided that a radiation detector in the distributed sensor network has detected an anomaly, the spectral information collected can then be utilized at other sensor locations. To increase detection sensitivity a detection algorithm can be operated on a specific area, or areas, of the collected energy spectrum, referred to as a region-of-interest. The objective of this strategy is to maximize

Table 4: Spectral comparison results for L^1 and L^2 norm methods.

	Spectra Type 1	**Spectra Type 2**	**Spectra Type 3**
L^1 Norm	671.0	616.5	0.15
L^2 Norm	60.0	48.2	0.01

Table 5: Spectral comparison results for Chi-squared, K-S Test and Cosine Similarity.

	Spectra Type 1	**Spectra Type 2**	**Spectra Type 3**
Two-Sample χ^2 Test	344.5	0.57	0.02
K-S Test (p-value)	0.18 (3.3E-15)	0.10 (1.4E-5)	0.03 (0.69)
Cosine Similarity	0.75	0.98	0.98

the signal-to-noise ratio by using the previously collected anomalous spectrum. For instance, if a known source distribution is primarily in a low-energy region, a detection algorithm that analyzes the entire spectrum (i.e. full gross counts) would add unwanted noise by summing the energies bins above the low-energy region. To minimize this noise contribution to the detection algorithm and thus increasing the signal-to-noise ratio, a detection can be applied only to the low energy region.

In [154], a relatively heuristic approach is used to accomplish this task. This approach returns a list of bins that maximize the signal-to-nose ratio, which could be utilized at down-stream sensor. This is accomplished by the following steps:

1. Combine spectra during anomalous event and divide by the total live time to time-normalize,
2. Calculate the mean background spectra by combining spectra prior to anomalous event and divide the total live time of the combined spectra to time-normalize,
3. Calculate the signal-to-noise ratio vector by dividing the source spectra by the square root of the sum of the source and background spectra,
4. Sort the signal-to-noise from greatest to least,
5. Step through the sorted vector and sum each element until the summation begins to decrease, and.
6. Return the bin indexes that maximize signal-to-noise ratio.

This method can be expanded to supply a region- or regions-of-interest within the energy spectra for increased detection sensitivity, rather than solely the individual bins. The returned bins are identical to the positive bins return from the background-subtracted spectra previously calculated in Figure 38. The regions-of-interest would therefore appear similar to the highlighted areas in

Figure 46. A smoothing filter, such as the methods discussed previously in this work, can be applied to decrease the sparsity and provide a more realistic representation of the source distribution. In this particular example, the filtered spectrum is continuous with no zero-valued bins resulting in a single region-of-interest, shown in Figure 47.

In Figure 47, the region-of-interest after the spectra are filtered results in a single region-of-interest is constructed. This may not always be the case, where multiple regions-of-interest may result even after the spectra are filtered due to low counting statistics or noise in the data. A more physical approach should be utilized where a single region-of-interest is guaranteed by selecting all bins up to the highest bin in the background-subtracted, filtered anomalous spectra. This allows for a more realistic region-of-interest as scattering interactions (e.g., Compton scattering) of the photons emitted from the radioactive source are typically dominant. Thus, in this work, a region-of-interest below the highest energy bin anomaly is used.

4.2.1: Multiple Gross Count Methods

With a region-of-interest selected, the conventional gross counts detection method can be utilized on the data within the region-of-interest. Caution should be exercised when selecting the appropriate decision threshold as now two detection algorithms are operated simultaneously (i.e. full spectrum gross counts and region-of-interest gross counts). Each detection method will have distinct true-positive and false-positive rates, thus naively selecting a decision threshold can potentially increase the detection system's false alarm rate. To maintain a constant false-alarm rate, each decision threshold must be scaled. One method to accomplish this is employ the Dunn-Šidák correction [154]. The Dunn-Šidák correction assumes each test, or detection method, is independent and error rates (α_n) for each method to meet the specified total error rate (α_T) is unknown. First, to derive the correction formula, the probability that at least one of the methods

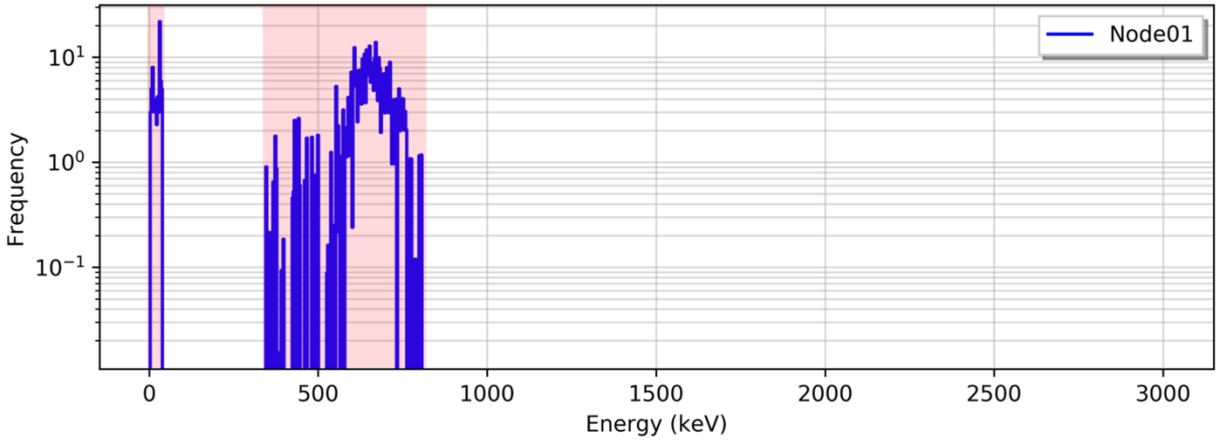


Figure 46: Background-subtracted spectra with highlighted regions-of-interest using heuristic approach.

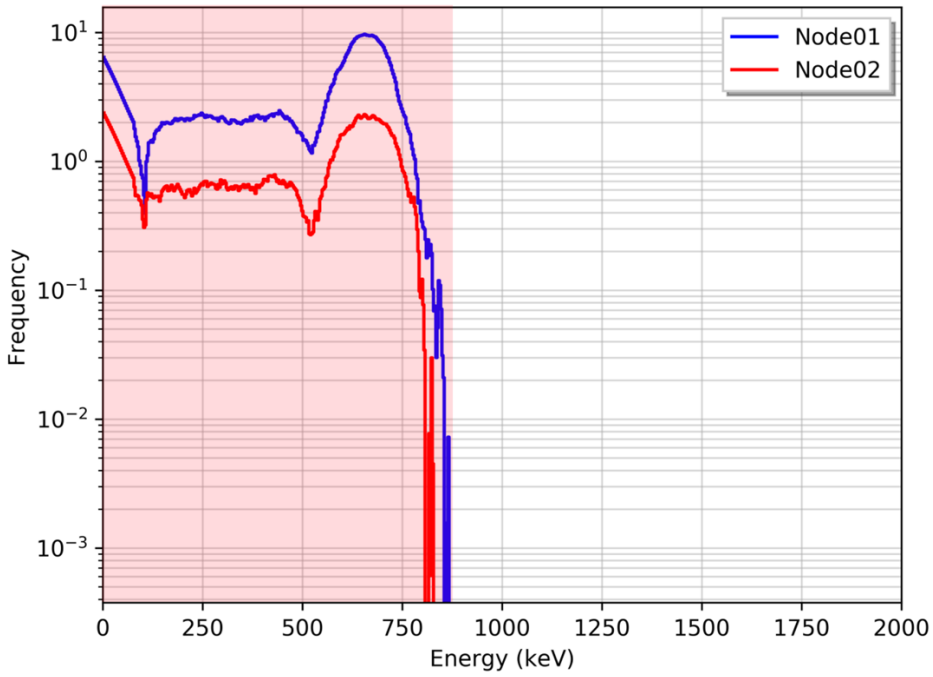


Figure 47: Filtered spectra with highlighted region-of-interest.

is successful given α_n is equal to one minus the probability that none of the methods are successful. Assuming the independence between detection methods, the probability that none of the methods are successful is equal to the product of the probabilities that each of method is not successful. This is illustrated by the following equation $1 - \alpha_T = (1 - \alpha_n)^n$ [154]. Subsequently solving for α_n provides the error rate to be employed for n detection methods, $\alpha_n = 1 - (1 - \alpha_T)^{\frac{1}{n}}$. For instance, if a confidence value of 95% and, thus, an error rate of 0.05 is specified for a detection system, then a confidence value 97.5% with an associated error range of 0.025 is required for a two-detection method approach. This approach will be utilized when using the conventional, full spectrum gross counting and the region-of-interest gross counting methods concurrently.

4.2.2: Multiple-linear Regression Method

An alternative approach employs a regression model to estimate the count rate within a specific region-of-interest using the count rate outside of the region-of-interest. The regression model attempts to learn the expected count rate for the region-of-interest and if the observed count rate is greater than the expected count rate, a radioactive source is most likely present. An extension of the ordinary least-squares regression method is utilized that uses multiple independent, explanatory variables (i.e. out-of-region spectrum bins) to predict the multiple dependent, response variables (i.e. in-region spectrum bins) via a linear model. This type of model attempts to create a linear relationship to approximate the dependent variables provided a set of independent variables. The model is calculated by first forming a linear equation where each independent variable (x) is weighted by a coefficient (β) added to an axis-intercept parameter (β_0) to predict a dependent variable (y). As it is near-impossible, if not impossible, to create a perfect regression model, some amount of error between the predict value and the

observed will occur. The multiple-linear regression model incorporates an error term (ϵ) to account for the error and is assumed to be a zero-mean, normally distributed value. The general formula for the multiple-linear regression model for n-independent, predictor parameters is provided as follows: $\boldsymbol{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \boldsymbol{\epsilon}$ and thus, $\boldsymbol{y} = \beta_0 + (\sum_1^n \beta_n x_n) + \boldsymbol{\epsilon}$**.** Of note, more complex regression models can be used when hypothesized to better represent the data, such as higher power orders on independent, predictor parameters or interactions between the independent, predictor parameters (e.g., multiplication). Regardless of the complex model chosen, these models will remain a linear regression model as the parameters are expressed a linear combination of the independent, predictor parameters.

The best-fit line for the observed data ($x_1, x_2, \dots, x_n$) is calculated by minimizing a cost function, typically mean squared error or residual sum of squares (RSS), of the deviations from each observed element in the data to the best-fit line. Minimizing the error is similar to maximizing the likelihood of $L(x|\beta_0, \beta_1, \dots, \beta_n)$, where both provide optimal coefficients for the regression model. This is a common technique with a plethora of literature discussing the vast assortment of use cases along with the full derivation of the equations utilized during minimization. In the following paragraphs, a shortened derivation of the methods utilized in this work is provided and the author suggests reviewing others works that provide additional information regarding the assumptions and reasoning behind the methods [156].

Using a sum of the squared residuals method on a set of data, the regression model can be moved into a matrix form where linear algebra methodologies can be employed to efficiently calculate the optimal coefficients. This is accomplished by first taking the sum of the squared residuals over the observed data (i.e. x, y) and the predicted data (i.e. $\hat{y}, \beta$), $RSS = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y_i})^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij}\right)^2$. The formula can be conveniently formulated

into a matrix form by first assigning the following matrices. Note that in this report, a conventional labelling method is employed where bold-faced lowercase letters denote vectors and the bold-face uppercase letters indicate matrices; $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$, $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$, and $\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$. Using the created vectors and matrices, the linear regression model can now be constructed into a more compact notation $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Recall that the goal is to minimize the residual error ($\boldsymbol{\epsilon}$), which in linear algebra terminology is the same as obtaining the coefficient vector ($\boldsymbol{\beta}$) that minimizes the following the equation $\sum_{i=1}^{n}(y_i - \hat{y_i})^2 = \boldsymbol{\epsilon}^T\boldsymbol{\epsilon} = (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})$ and, subsequently expanding the operations, $(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}) = \boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\boldsymbol{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$. The latter equation takes advantage of the fact that a transpose of a scalar is the scalar, resulting in double the value in this case (i.e. $\boldsymbol{y}^T\mathbf{X}\boldsymbol{\beta}$ is equal to $\boldsymbol{\beta}^T\mathbf{X}^T\boldsymbol{y}$). Now the partial derivative with respect to the coefficient vector provides an equation in matrix form that can be utilized for calculating the optimal coefficient values. The partial derivative is then set equal to zero and the coefficient can be separated from the other terms, as shown in the following derivations $\frac{\partial(\epsilon^T\epsilon)}{\partial\beta} = -2\mathbf{X}^T\boldsymbol{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = 0$ and, thus, $\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\boldsymbol{y}$. This derivation only exists if the inverse exists (i.e. columns in $\mathbf{X}$ are linearly independent or if more independent parameters than observations exists, $n > k$), where if the opposite if true and the inverse does not exit, then the solution may not unique. This solution provides a method for obtaining the optimal coefficients in the original regression equation solely based on the observed independent data ($\mathbf{X}$) and the observed dependent data ($\boldsymbol{y}$). The derived equation it typically used to solve for the optimal coefficient vector using observed data and subsequently insert into the original

regression equation to predict the dependent values. A hat, positioned above the dependent parameter ($\widehat{\boldsymbol{y}}$), is utilized in this context to denote an estimator or an estimated parameter. Eliminating the residual error term from the original equation as the optimal solution attempted to provide a near-zero sum of squares, new dependent values can be approximated $\widehat{\boldsymbol{y}} = \mathbf{X}\boldsymbol{\beta} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\boldsymbol{y}$. A matrix referred to as the hat-matrix (H) is often utilized to shorten the derived equation by incorporating all the independent vectors into a single parameter that maps the observed data to the unobserved data $H = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T$ and $\widehat{\boldsymbol{y}} = H\boldsymbol{y}$.

The variance of the regression model is important to measure as the accuracy of the model and the confidence of a decision threshold depend on the variance. A relatively common approach to calculate the model's error, or standard deviation of the model, is the mean squared error (MSE). The MSE for the calculated regression hyperplane is equal to the sum of the squared errors between the predicted and observed dependent values divided by the degrees of freedom (d) minus one. The degrees of freedom value is calculated by subtracting the number of samples (n) by the number of independent predictors (k); that is, $MSE = \frac{\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}}{d-1} = \frac{\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2}{n-k-1} = \hat{\sigma}^2_{MSE}$ and $\hat{\sigma}_{MSE} = \sqrt{\hat{\sigma}^2_{MSE}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2}{n-k-1}}$. This can also be referred to as the root mean-squared error or square root of the mean square error. Note that the standard deviation equation requires more observations than the number of independent variables, as the denominator would become negative resulting in an imaginary number for the standard deviation. Recall that this is also required for the inverse to exist. The standard deviation is used to weight the difference between the expected and observed in-region counts, as shown in Figure 48.

At this point, the above multiple-linear regression model could be sufficient in providing an estimated count rate in the region-of-interest by using the count rates outside the region-of-

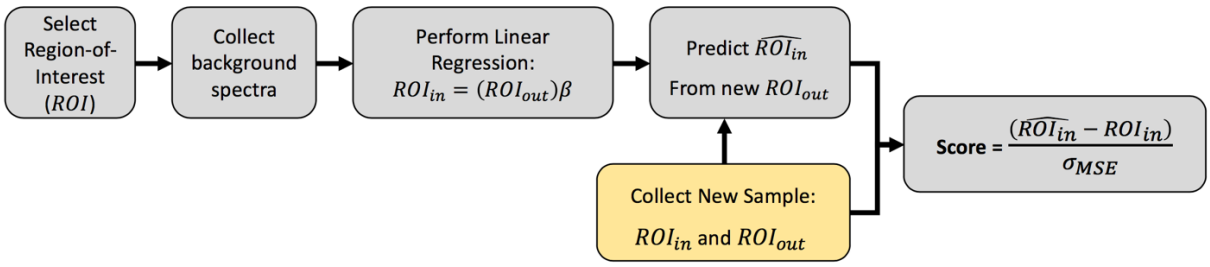


Figure 48: Multiple-linear regression detection methodology using region-of-interest approach.

interest. This model will be utilized later using synthetic and real spectral data, but an alternative approach that expands upon the regression method is first derived.

To avoid overfitting the data and provide a more robust regression model, a more recent method called Tikhonov regularization can be applied. Tikhonov regularization, which is more commonly referred to as ridge regression or L^2 regularization, expands upon the previously derived multiple-linear regression model and overcomes the primary issues previously discussed; namely, instances where there exist more predictor parameters than observed parameters and the inability weight parameters based on importance. Furthermore, ridge regression has been traditionally employed when multicollinearity is present, meaning the independent variables exhibit a correlation amongst themselves. In the case of analyzing collected data in radiation spectra, correlations can exist between bins as the underlying physical properties that govern the frequency can be strongly correlated.

As shown previously, the multiple-linear regression model adjusts the coefficient vector ($\boldsymbol{\beta}$) based on the observed data, referred to as training data in the machine learning context. As noise is generally inherent to all observed data, the coefficients will optimize to values that model the noise effectively. This is where regularization methods can benefit. Ridge regression accomplishes this task by using a ridge estimator, or tuning parameter, that increases the bias to improve the variance of the model, such as changing the slope of a line in a two-dimensional space. Admittedly, this can result in lower performance when training the regression model on the observed data but potentially perform well on both observed and predicted data while not overfitting. This concept is illustrated in Figure 49. This tuning parameter (λ) decides the extent to which the flexibility of the regression model is penalized. This is accomplished in the

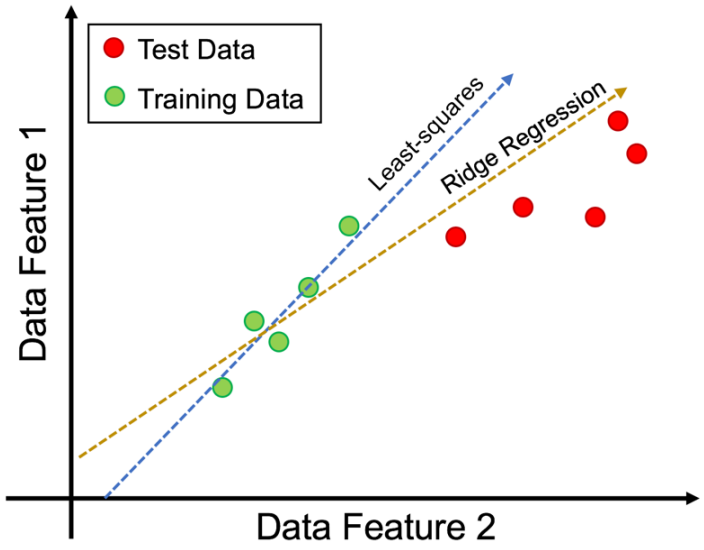


Figure 49: Illustration comparing least-squares and ridge regression on arbitrary data.

minimization step by adding a penalty term to the residual error, as shown in $RSS_{Ridge} = RSS + \lambda \sum_{j=1}^{k} \beta_j^2$ and $RSS_{Ridge} = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij}\right)^2 + \lambda \sum_{j=1}^{k} \beta_j^2$. The tuning parameter is constrained to be equal to or greater than zero, where a zero-valued tuning parameter results in the identical residual sum of the squares (RSS) as least-squares regression. Following the derivation steps provided previously for the multiple-linear regression, the equation can be placed in matrix form and the ridge coefficient vector ($\boldsymbol{\beta}_R$) can be isolated after performing the partial derivative. This results in the following equation, where tuning parameter is multiplied by the identity matrix ($\mathbf{I}$) creating a diagonal *ridge* that is added to the correlation matrix ($\mathbf{X}^T\mathbf{X}$) via $\boldsymbol{\beta}_R = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\,\mathbf{X}^T\boldsymbol{y}$. The additional term in the ridge sum of the squares (RSS_{Ridge}) containing the summation over the entire coefficient vector squared is commonly referred to as the shrinkage term and the tuning parameter controls the amount of shrinkage. A large tuning parameter forces the coefficient estimates to approach zero and, thus, reducing the model variance. This increase in bias can be beneficial in reducing the variance to a point but can begin underfitting the data and lose important properties in the data. This is illustrated in Figure 50 using arbitrary data. For an amount of shrinkage, which is identical to the tuning parameter value, an optimal value can is obtained at the minimum of the prediction error. In general, the bias increases as the tuning parameter increases and the variance decreases as the tuning parameter increases (i.e. bias-variance tradeoff). Standardization of the data is required in ridge regression in order to even the amount of change, or shrinkage, for each coefficient. For instance, if the observed data were on vastly different scales, the shrinkage would be biased as their contributions to the penalty term would differ because the penalized term is the sum of the squares of the coefficients. Scaling of the data decreases the need for large values in the coefficient vector ($\boldsymbol{\beta}_R$), resulting in a smaller, optimal tuning parameter. A conventional solution

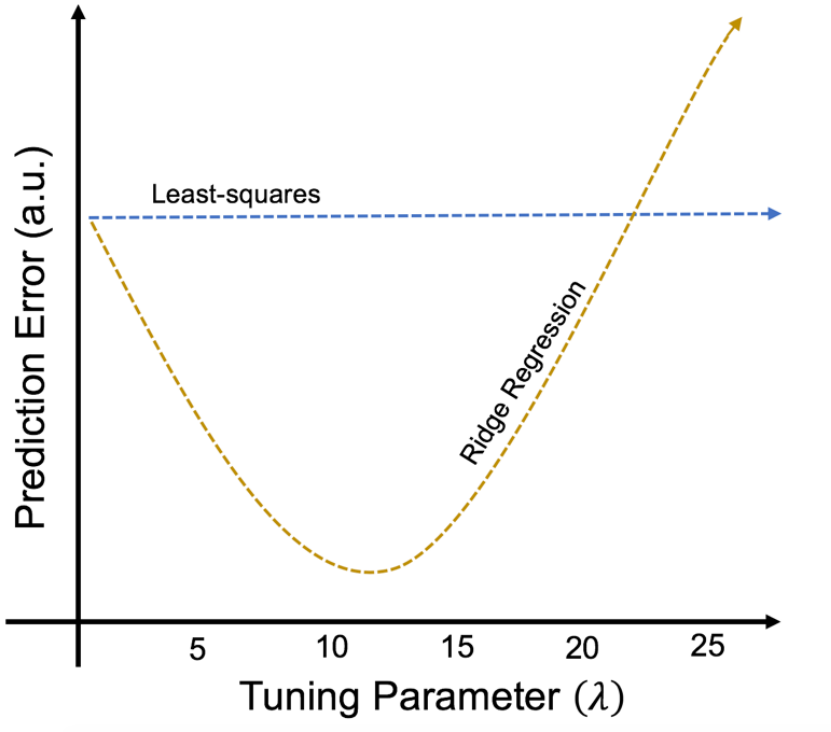


Figure 50: Example illustration of prediction error for least squares and ridge regression methods.

is to mean-center and scale the data to have a unit variance (i.e. variance equal to one) prior to performing ridge regression. Thus, both the independent ($\mathbf{X}$) and dependent ($\boldsymbol{y}$) data are scaled through subtraction of their means and divided by their standard deviation providing a variance of one.

It is worth noting that when analyzing sparse data where many zero-valued features exist, ridge regression can fail to provide accurate predictions for said features as the coefficients calculated in the ridge regression model cannot be exactly zero, but rather very close zero. Generally, this issue is addressed by a process called variable selection and requires prior grooming of the data to select only relevant data features to analyze [161]. This issue can also decrease the predictive accuracy of the ordinary least squares regression method quite substantially. In this case, radiation spectra can often be sparse with a moderate to high quantity of empty bins. This is illustrated in Figure 51 providing a histogram of count rates for a detector in the ORNL sensor network for a 0.0 to 3.0 MeV energy range.

This hindrance is decreased when using the region-of-interest estimation approach, where the out-of-region frequency estimations are typically nonzero. This does not completely alleviate the issue, as scenarios do exist where zero-valued regions occur and are subsequently used to predict the in-region frequency. For instance, if the region-of-interest consumes the majority of spectrum (e.g., 90% of the bins), then the region outside the region-of-interest could have a zero-valued frequency (e.g., high photon energy regions due to low photon flux or minute interaction cross sections). The author thus suggests utilizing a threshold for using the region-of-interest regression methods, where the out-of-region frequency must be equal to or greater than the threshold, say 20 counts per second. With a mean count rate equal to or greater than 20, a normal

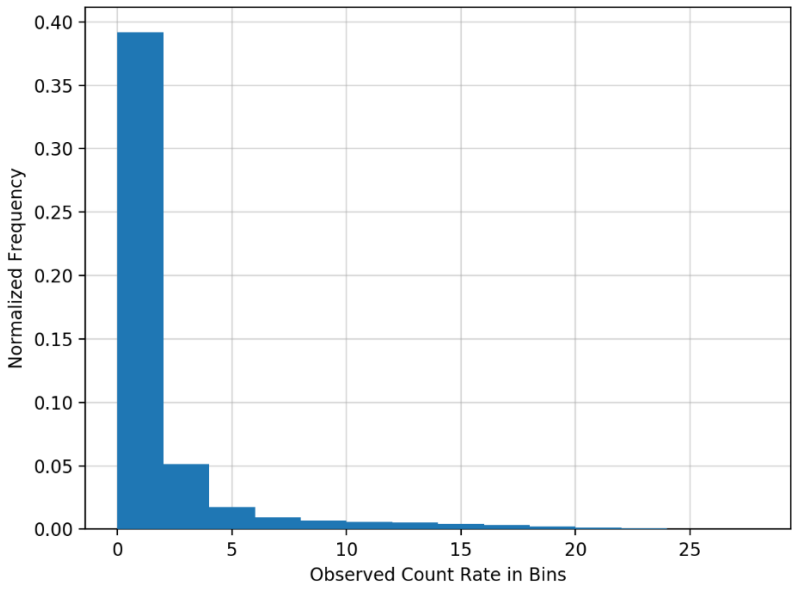


Figure 51: Histogram of count rates per bin for a deployed radiation detector.

approximation can be made confidently for a Poisson process and small values are probabilistically not likely.

Moving forward, the overall approach of using ridge regression in this context, shown in Figure 52, is similar to the multiple-linear regression approach with an additional step of optimizing the tuning parameter in the penalty term of the regression model. Selecting the optimal tuning parameter (λ) is a rather controversial undertaking and several methods can be utilized. In [159], Hoerl suggest a graphical approach by plotting each regression coefficient as a function of the tuning parameter value. Using this graph, referred to as a ridge trace, an analyst has the ability of selecting the smallest tuning parameter where the regression coefficients stabilize. The coefficients can vary greatly at low tuning parameter values but selecting the smallest value when the coefficients stabilize avoids introducing unwarranted bias to the model [159, 160]. To provide context, an example ridge trace plot is provided in Figure 53 for artificially created data. From Figure 53, the ten coefficients are initially spread out and begin to converge as the tuning parameter increases. A tuning parameter value can be selected by choosing a value prior to the convergence, say 5E-6 in this example, where the coefficients appear relatively stabilized.

Unfortunately, using the ridge trace graph to select the tuning parameter has no objective basis and is prone to user bias. In more recent years, other methods have been proposed for selecting the optimal tuning parameter value, which include using techniques such as: Akaike's Information Criterion, Bayesian Information Criterion, and, most notably, cross-validation [161, 163, 164]. These techniques assist in the model selection phase when analyzing data but are not addressed in this work. The reason being is that by using the region-of-interest approach for the

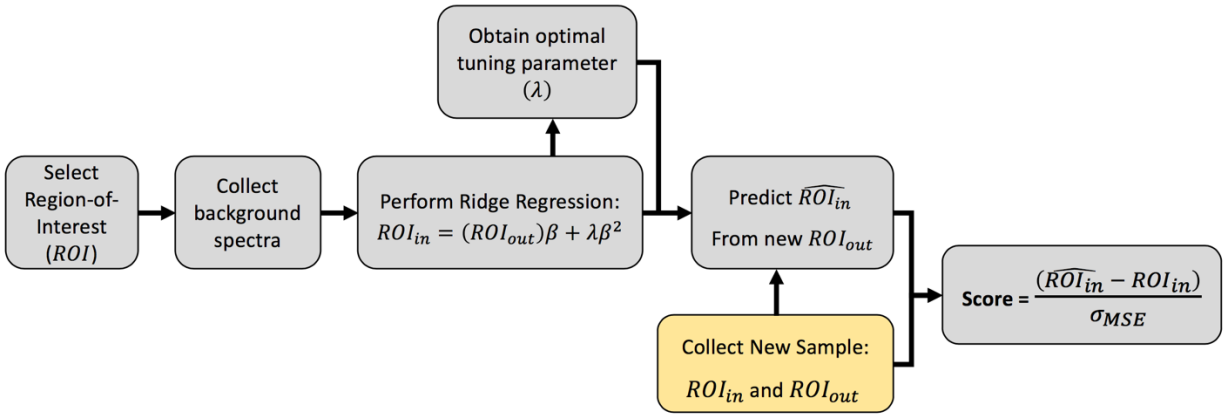


Figure 52: Ridge regression detection methodology using region-of-interest approach.

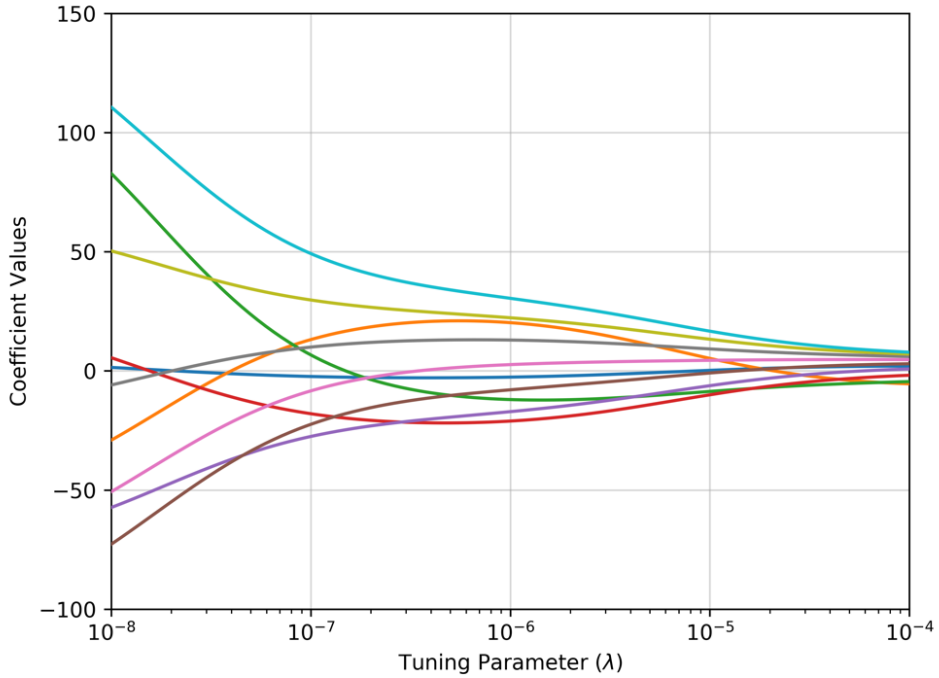


Figure 53: Example ridge trace plot for multiple coefficients using arbitrary data.

energy spectra, there only exists a single coefficient in the coefficient vector ($\boldsymbol{\beta}$). Thus, a more simplistic approach is to select the tuning parameter that meets a specific stabilization rate.

To assess the relationship between the ridge coefficient and the tuning parameter on real-world radiation data, a ridge trace is provided in Figure 54 for each radiation detector in the ORNL sensor network using the region-of-interest highlighted in the previous example spectra, shown in Figure 47 (i.e. approximately 0.0 to 0.8 MeV). Figure 54 shows the ridge trace using 24-hours of radiation data with no anomalous signals and illustrates the previous claim that a large tuning parameter moves the coefficient estimates towards zero. The coefficient values remain relatively flat below a tuning value of approximately 10, which is an interesting outcome as a small tuning parameter will decrease the penalty term in the ridge regression model, thus approaching a linear regression model. This is shown in Figure 55 that provides the mean squared error for the ridge and multiple-linear (i.e. referred to as least-squares in Figure 55) regression models for each node on the ORNL sensor network.

Using the mean square error to compare the two approaches, the ridge regression is concluded to not provide any benefit compared to the multiple-linear regression model using the region-of-interest approach presented in this work. A receiver operator characteristic curve is utilized to compare the detection sensitivity the proposed methods provide over the full spectrum, gross count detection method (see Figure 57). Synthetic data was created using the source and background distributions provided previously in Figure 36 to compare the methods. Using a one second timing resolution, the source distribution is introduced for 4 seconds every 60 seconds with a constant background source term. This is illustrated in Figure 56 containing the gross count rate, the count rate within the region-of-interest, and the count rate outside the region-of-interest.

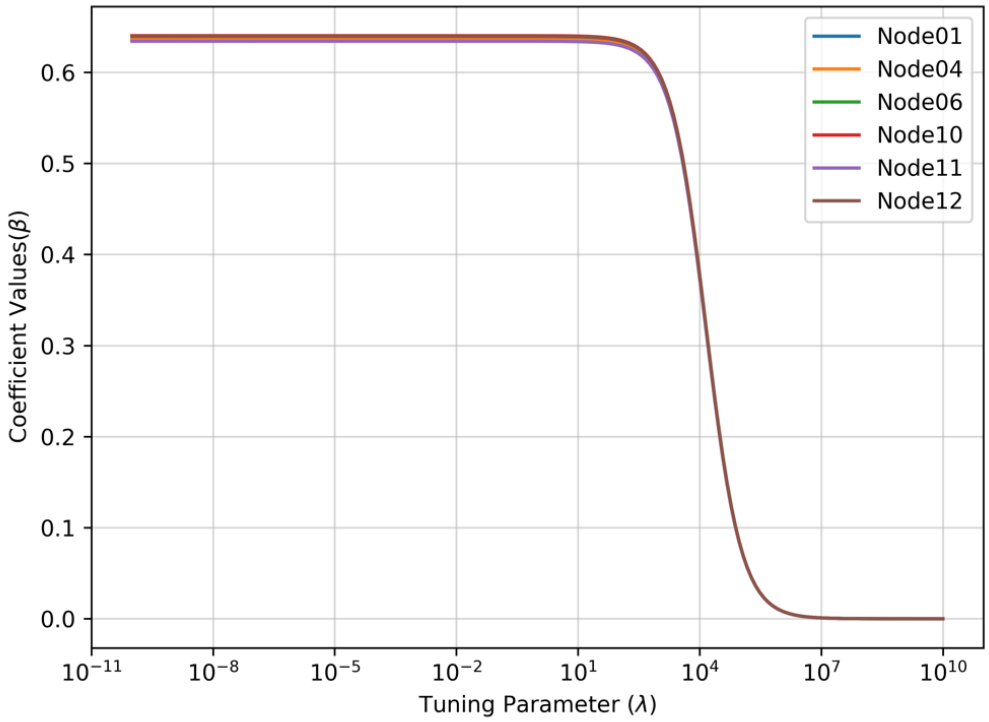


Figure 54: Ridge trace for radiation detectors deployed at ORNL.

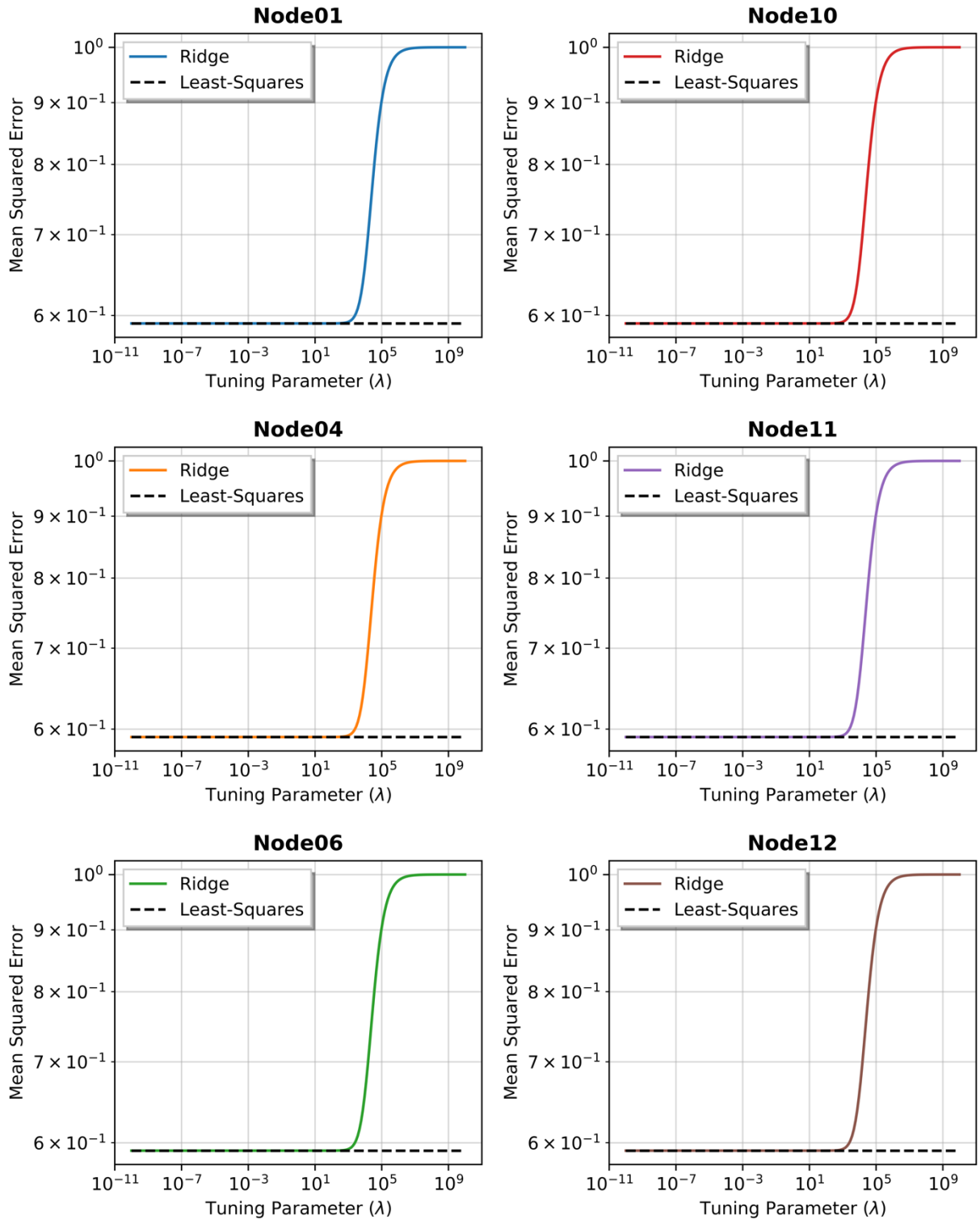


Figure 55: Mean squared error for ridge and least squares regression models on each ORNL sensor node.

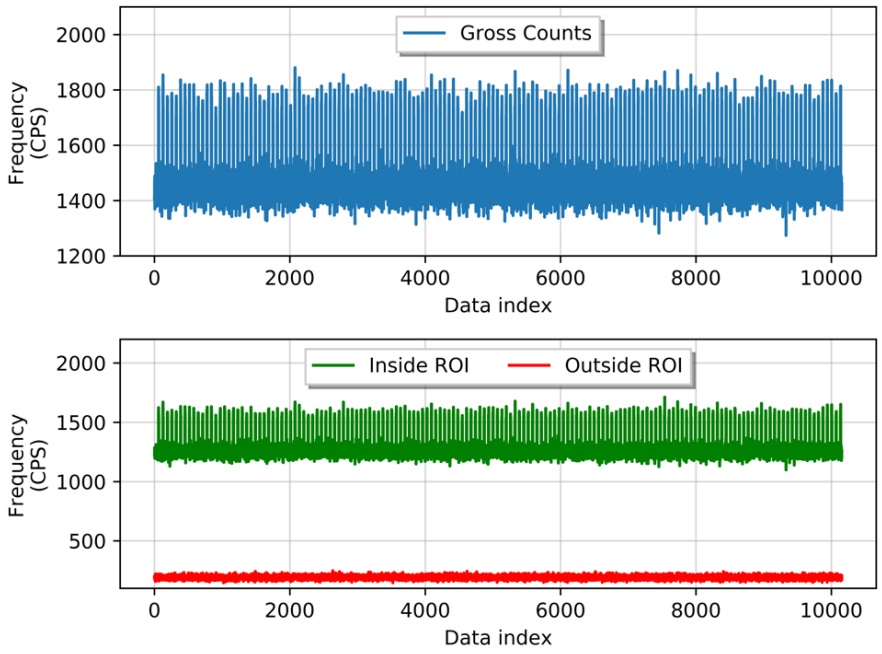


Figure 56: Count rate of synthetic data for comparing detection methods.

(Top) Gross counts of the synthetic data. (Bottom) Count rate within the region-of-interest (green) and outside the region-of-interest (red).

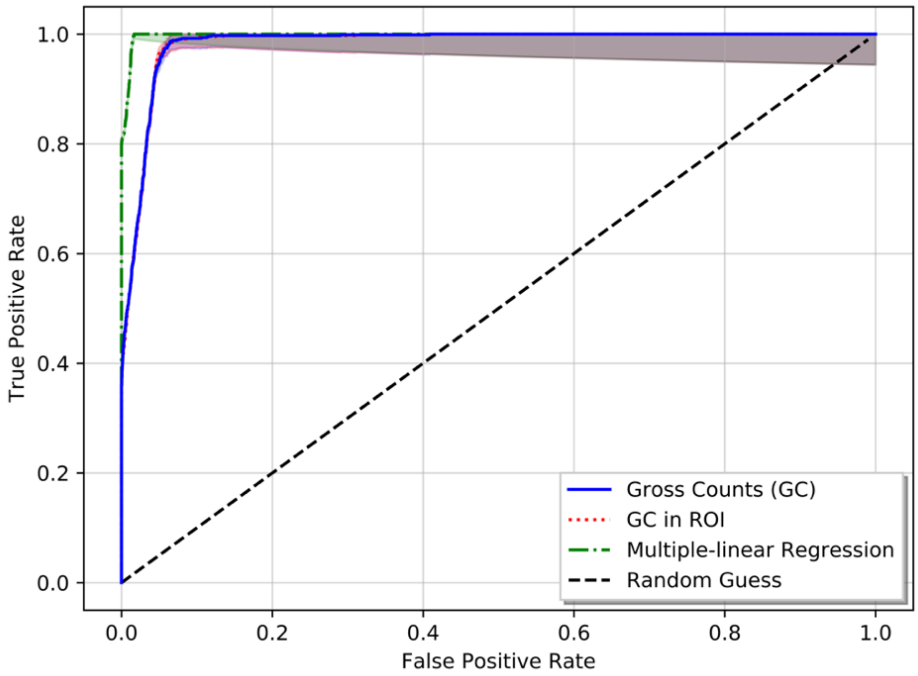


Figure 57: Receiver operator characteristic curve comparing gross counts and least-squares methods.

The gross counts method and the linear squares approach utilize a rolling-window approach to analyze time series data to account for temporal changes. Two rolling-windows are employed (i.e. one for the background and another for the foreground), where the width of each window is selected prior to executing the algorithm. In this analysis, a 10 second background window and a 2 second foreground window are used with a 4 second separation between the two windows. The windows are separate to account for slowly changing anomalies, where side-by-side rolling-windows would have less sensitivity as the slowly increasing count rate would be contributing the background, or expected, distribution. The resulting receiver operator characteristic curves are provided in Figure 57. Wilson score intervals were utilized to provide a confidence interval at a 95% confidence level with an error rate (α) of 0.05 and are represented by the shaded regions in Figure 57. The intervals are rather large at low decision threshold values (i.e. right side of curve) due to the high number of false positive classifications, increasing the uncertainty. The multiple-linear regression provides the better prediction method, which is illustrated by its high true positive rate at low false positive rate values. Both the region-of-interest gross count method and the conventional gross count method maintain a competitive prediction capability with minute differences between the two.

Operating an additional detection algorithm may provide increased detection capability or sensitivity, but at the cost of combining multiple error rates. Recall that to account for this change and to maintain a specified error rate, the Dunn-Šidák correction can be utilized. If a total error rate of 0.05 is specified, the resulting error rate for a two-detection algorithm approach requires an error rate of 0.025 for each detection algorithm to obtain the desired 0.5 error rate. Using the updated failure rate, the methods are compared to the conventional gross counts (Option 1) method using the probability of detection metric. The results are provided in Table 6.

The probability of detection decreases when using the second option in Table 6, where both the conventional gross count and the region-of-interest gross count detection methods are used. The data on the receiver operator characteristic curve for both of these detection methods nearly perfectly overlaps, thus using the region-of-interest coupled with a more stringent confidence level results in lower detection probability or true positive rate compare to the conventional approach in Option 1. The multiple-linear regression model provides an increased sensitivity, which is reflected by the separation of the data receiver operator characteristic curve in the upper left corner. Even when utilized the Dunn-Šidák correction, the resulting probability of detection is increased from 0.95 to 0.99.

3.3: Analysis of Sensor Array Data

The methods previously discussed are extended to analyze several anomalous events observed on the ORNL sensor array. The raw outputs for the spectral comparison and region-of-interest detection methods are provided in this section and are further evaluated in the following section for a decision-making architecture. Therefore, this section solely supplies the reader with several the previous methods' outputs on real-world examples.

Each anomalous event begins by first detecting an anomaly at one of the sensors in the array via the gross count algorithm and, subsequently, calculating a region-of-interest after filtering the background-subtracted anomalous spectra. Three anomalous events are analyzed that traverse through the ORNL sensor array to assess the approaches discussed. The first two events (i.e. Event 1 and Event 2) follow identical paths starting at Node06 and the third event (i.e. Event 3) starts at Node 04 with all events exiting the area at Node01, as depicted in Figure 58 and Figure 59. Additionally, all three of the anomalous events occur due to transport of

Table 6: Probability of Detection for multiple detection options on synthetic data.

	Error Rate per Method	**Total Error Rate** (α)	**Probability of Detection**
Option 1: Full GC	0.05	0.05	0.95
Option 2: Full GC and ROI GC	0.025	0.05	0.69
Option 3: Full GC and ROI MLR	0.025	0.05	0.99

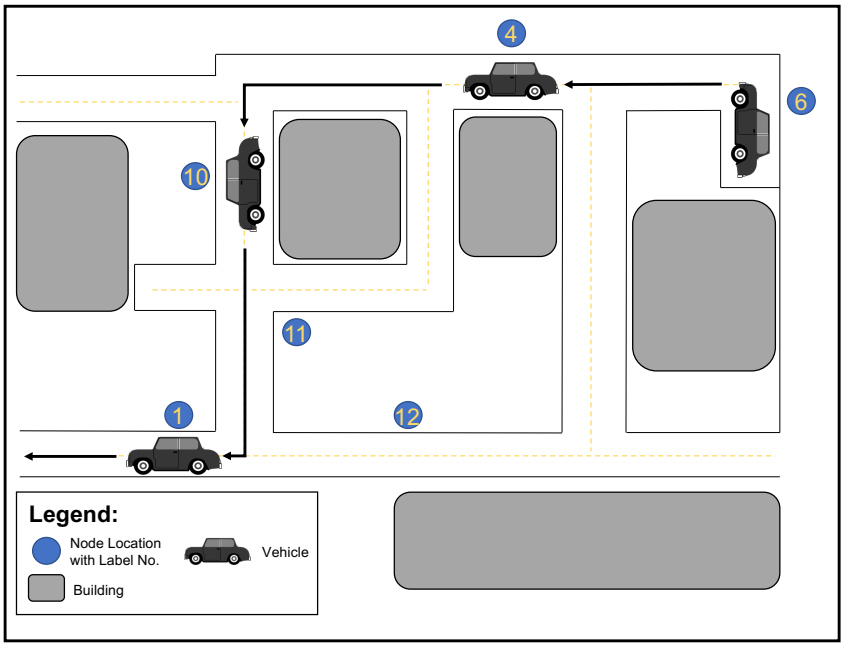


Figure 58: Illustration of vehicle carrying radioactive material through ORNL sensor array for Events 1 and 2.

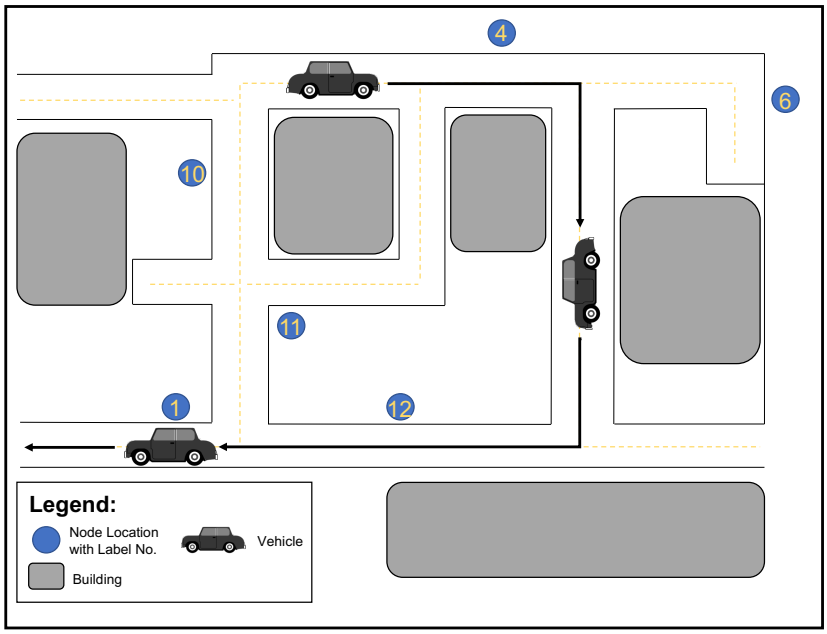


Figure 59: Illustration of path of a vehicle carrying radioactive material through ORNL sensor array for Event 3.

radiological material in a vehicle, thus short interactions are observed from the radiation sensors due to the moderate speed of the vehicles. The radioactive material moving through the ORNL sensor array is different for each event, resulting in unique observed count rates and spectral features. Each event is analyzed in order from Event 1 to Event 3 following the specific figures.

In Event 1, four nodes detect an anomaly due to the high increases in count rate during the time period of 13:09:15 to 13:10:50. The radioactive material traveling through the sensor array during Event 1 is a mixture of unknown radioisotopes with three prominent peaks. This is observed in Figure 61 showing the summed energy spectra during the anomalous event along with a one-minute summed background energy spectra prior to each detected anomaly.

Qualitatively comparing anomalous spectra observed at Node01, Node04, Node06, and Node10, high count rates are observed below approximately 1400 keV from the unknown radioactive material with three obscure peaks at approximately 760 keV, 1170 keV, and 1330 keV. Regardless of the exact radioisotope, or radioisotopes, in the material, the two approaches previously discussed of comparing anomalous spectra and increasing probability detection by operating an additional algorithm are utilized. The region-of-interest is first created on the initial anomalous, background-subtracted, filtered spectrum collected at Node06 shown in Figure 62. In this case, the filtering of the anomalous spectrum does not provide much benefit as the source strength and the total time integrated during the alarm are both relatively large, resulting in a dense anomalous spectrum rather than a sparse spectrum with multiple empty bins.

The Event 1 time-series of the decision metrics for both the conventional gross counts and the multiple-linear regression detection methods is shown in Figure 63. The decision metric values are normalized by mean subtraction and dividing by the standard deviation to give a zero-mean, unit variance distribution. This allows the two methods to be compared on similar scales.

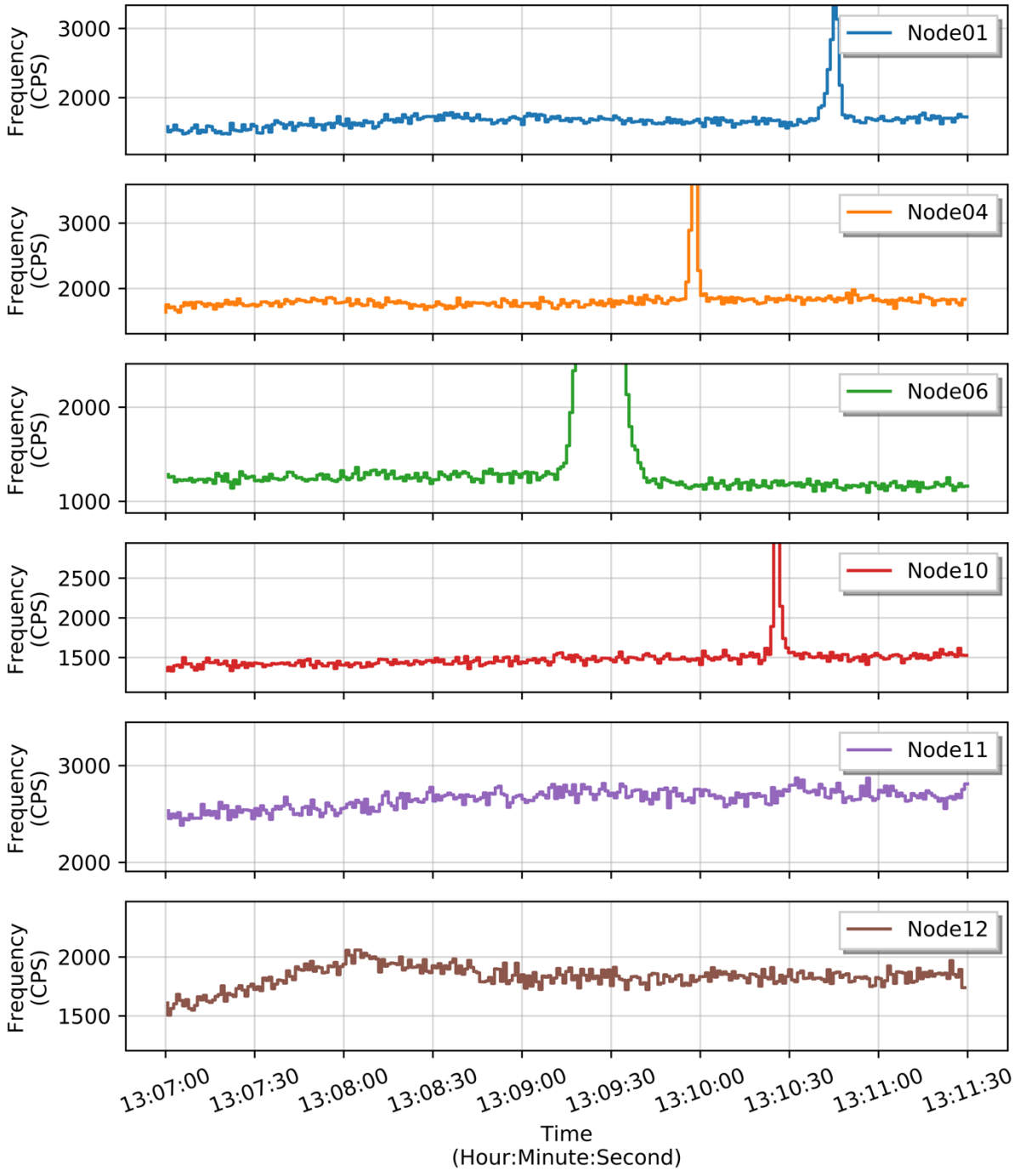


Figure 60: Event 1 time series of count rates observed on the ORNL sensor array.

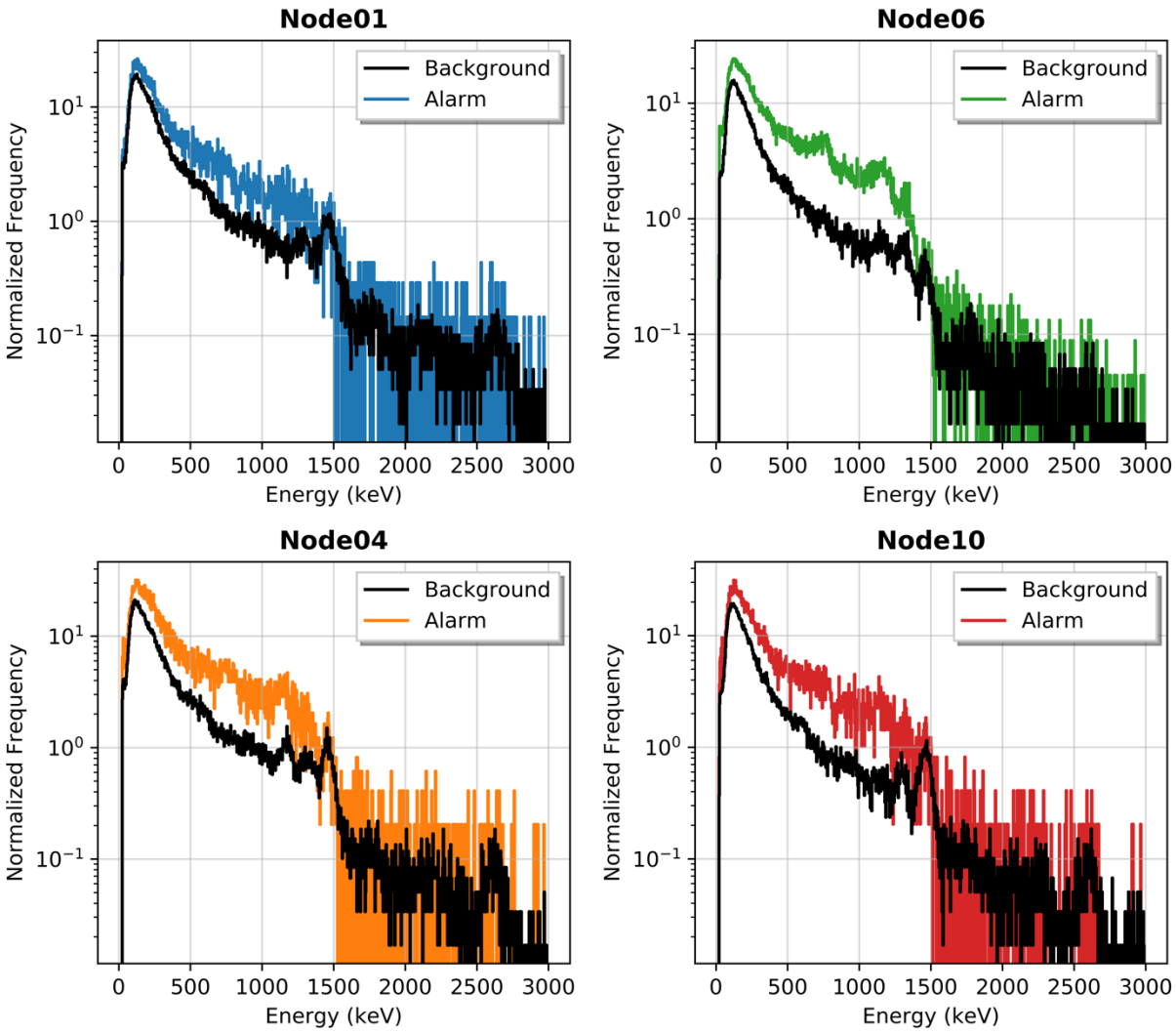


Figure 61: Event 1 alarm and background spectra collected at four nodes.

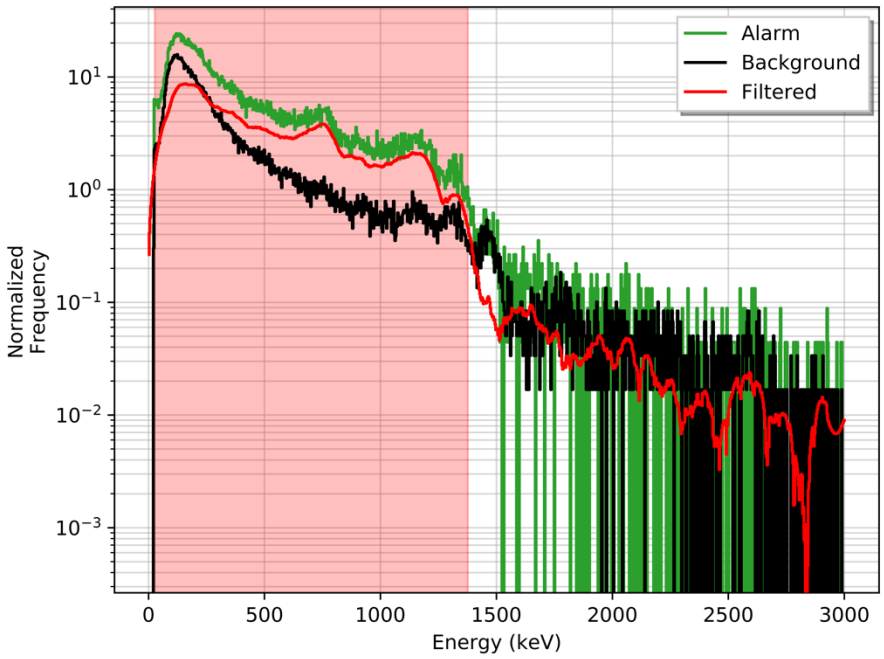


Figure 62: Event 1 region-of-interest from filtered anomalous spectrum at Node06.

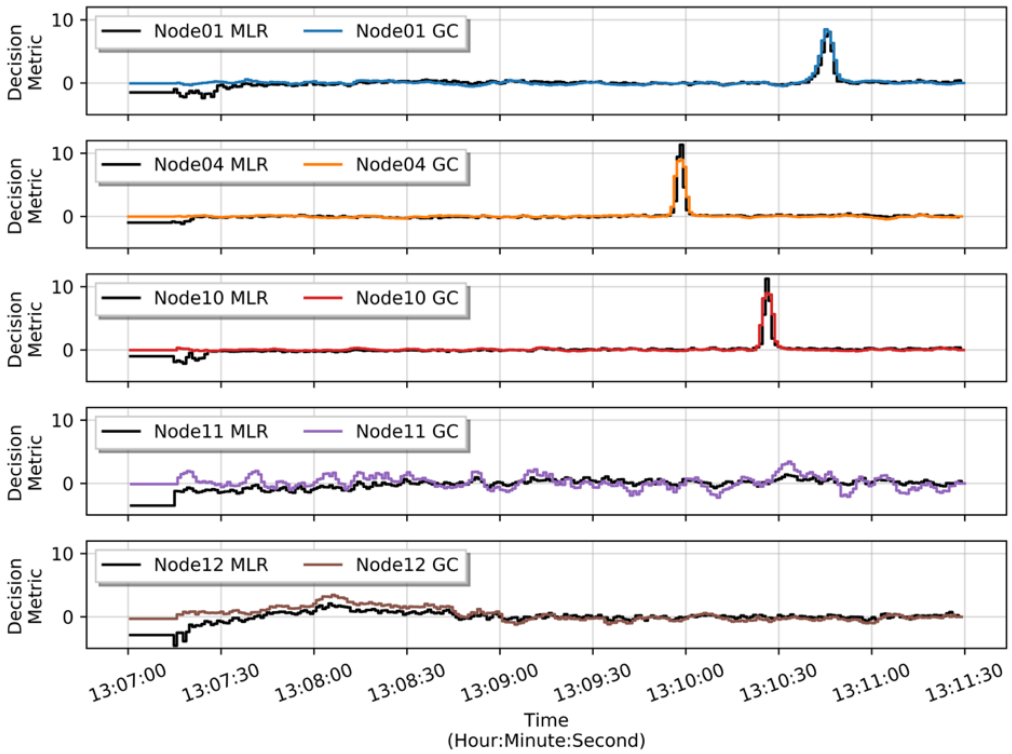


Figure 63: Event 1 decision metric time-series for remaining nodes.

Using a 95% and 97.5% confidence value for the single detection algorithm method (i.e. gross counts algorithm) and double detection algorithm method (i.e. multiple-linear regression and gross counts algorithms), respectively, the anomalies are correctly detected at Node01, Node04, and Node10. The multiple-linear regression method does provide a slightly higher probability of detection on Node04 and Node10 than the conventional gross count detection algorithm, but both methods in this case provides greater than 99.0% probability of detection due to the large count rate increase.

The combined spectra during the anomalous events at Node01, Node04, Node06, and Node10 are provided in Figure 64 after background-subtracting, smoothing, and applying the sum-to-unity normalization. Recall that the normalization is not required for the cosine similarity method but is utilized in this work to provide a more intuitive, qualitative comparison. Discrepancies between the filters are observed in high energy bins (i.e. > 1.5 MeV) which is primarily due to the low count rate in that region during the background-subtraction resulting poor statistics. This is compounded when the filtering method is applied resulting large disagreements between the spectra.

At lower energies, the peak previously identified around 760 keV, 1170 keV, and 1330 keV appear to align along with near identical intensities in the energy bins below 1400 keV. The cosine similarity method is utilized to compare the anomalous spectra collected from Node01, Node04, and Node10 with the initial anomalous spectra from Node06. Recall that the cosine similarity method provides a quantitative value from 0.0 to 1.0, where a value near 1.0 is represents strong similarity between the two spectra and is assumed that both spectra are samples from the same source distribution.

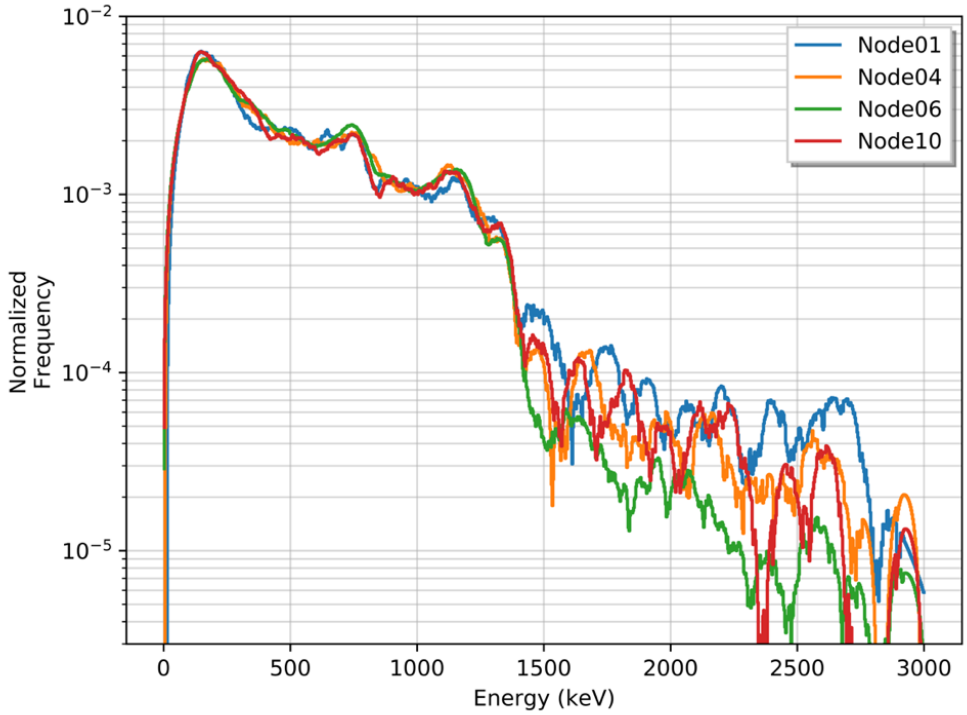


Figure 64: Event 1 anomalous spectra after background-subtracting, filtering, and normalizing.

The cosine similarity values for each collected spectrum Node01, Node04, and Node10 to the Node06 spectrum is provided in Figure 65 along with the cosine similarity values for the combined anomalous spectra provided in Table 7. The similarity values for the collected spectra in time-series are relatively high for background spectra, generally maintaining a value greater than 0.8 for Node01, Node04, and Node10 in Event 1.

In particular, Node04 in Event 1 observes a median cosine similarity value of 0.85 primarily due to the background spectra at the location is complex with energy peaks near the observed peaks in the anomalous spectra. This forces a more stringent threshold to be specified in order to confidently state the spectra collected is a sample from the same distribution as the combined anomalous spectrum from Node06. Of note, the time-series cosine-similarity values presented in Figure 65 are not background-subtracted, thus higher values is expected as there can exist large correlation values at lower energies, where both the background and anomalous spectra have similar shapes due to down scatter from higher photon energies. This issue only increases if no spectral features are present, such as an increase in counts across the entire energy spectrum which will be shown later. To reiterate, the analysis performed in this section is to introduce several scenarios observed on the ORNL sensor array and employ the down-selected methods on the real-world data. The output of each of the down-selected methods will be utilized in the actual decision-making framework to track the anomalies in the following sections.

Moving to the next anomalous scenario, Event 2 observes radioactive material initially being detected at Node06 then, similar to the path travelled in Event 1, travelling through the sensor array beside Node04, Node10, and Node11 before exiting the area at Node01. The time-series count rate for each node for Event 2 is provided in Figure 66. In this event, the radioactive material was initially positioned in the vicinity of Node06, which can be seen by the step

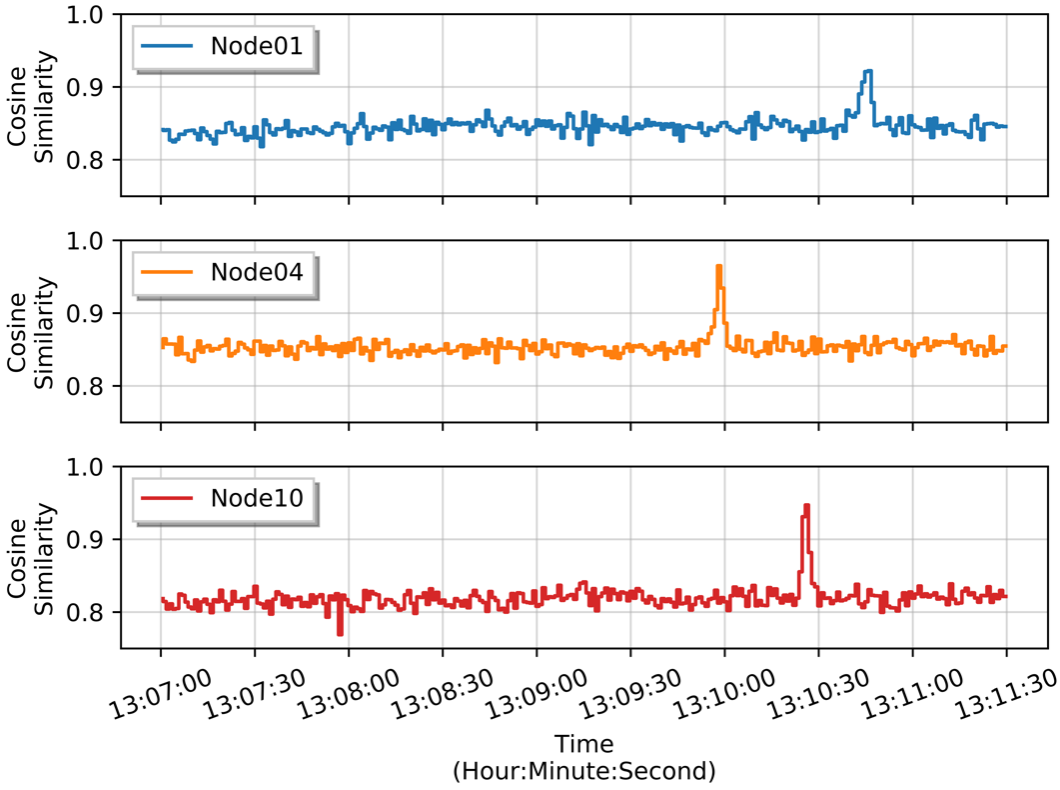


Figure 65: Event 1 cosine similarity time-series.

Table 7: Event 1 cosine similarity values for anomalous spectra.

Node	**Cosine Similarity**
Node01	0.993
Node04	0.998
Node10	0.996

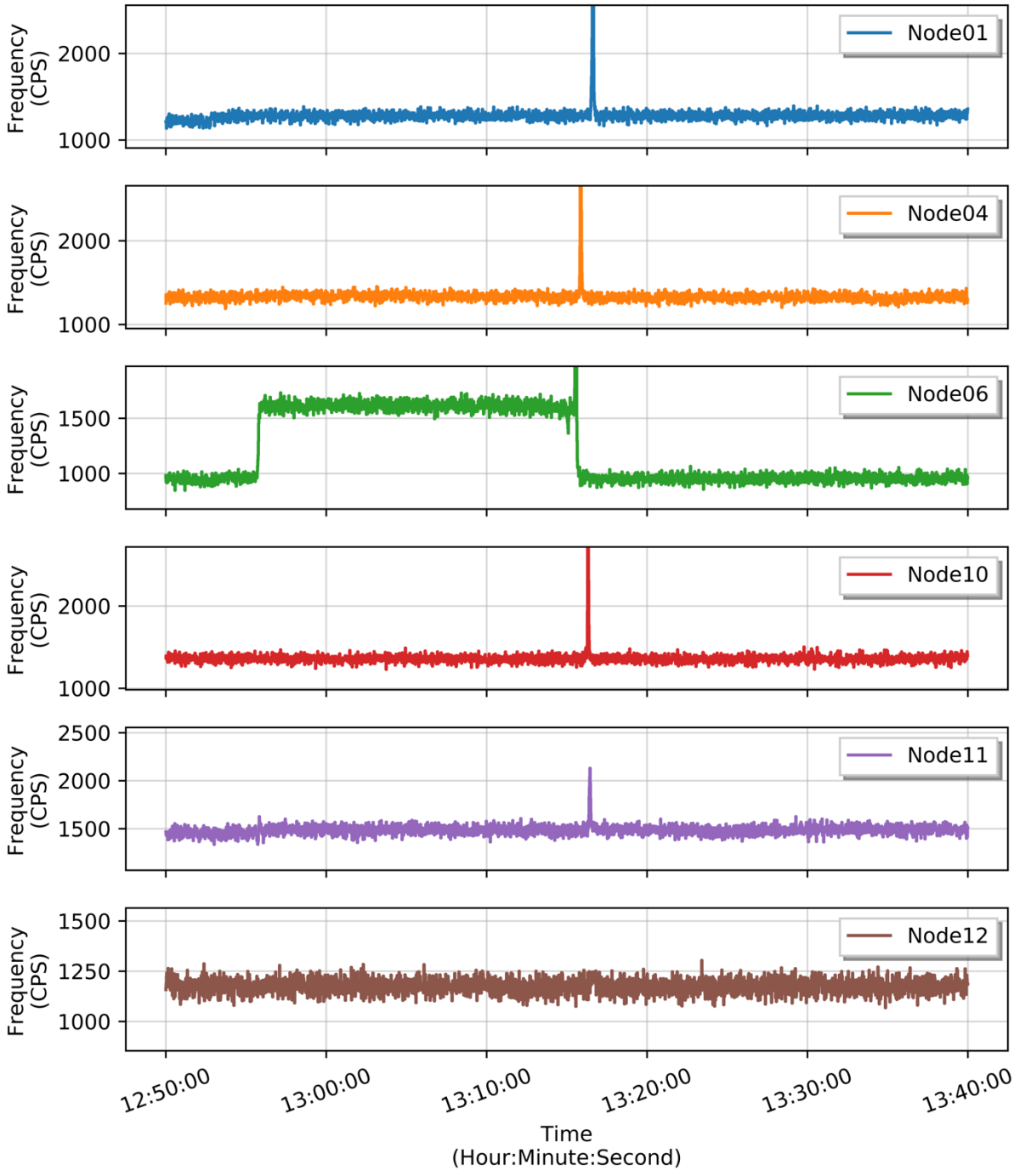


Figure 66: Event 2 time series of count rates observed on the ORNL sensor array.

function increase in count rate above the background at approximately 12:55:00 in Figure 66. Due to this long dwell time and high relative count rate, it is assumed that the alarm spectra should not be sparse and any spectral features inherent to the radioisotopes should be present.

The spectra are combined during the detected alarm period from 12:55:45 to 13:15:40 on Node06. The resulting anomalous spectra is provided in Figure 67 along with the calculated the region-of-interest from 0.0 to 1.39 MeV. Worth noting, applying the filter to the spectrum provides limited benefit in this particular case, as the spectra is not sparse due to the prolonged exposure time coupled with a moderate source activity. From the initial, observed anomalous spectra from Node06, two prominent energy peaks are observed around 1770 keV and 1330 keV, most likely resulting from ^{60}Co in the radioactive material. These peaks and their associated down scatter result in a calculated region-of-interest from 0.0 to 1400 keV. Using this region-of-interest, the for the single detection method and double detection method were both employed to compare the techniques. The decision metric values are provided in for other nodes in the ORNL sensor array. The decision metrics are normalized by mean subtracting and dividing by the standard deviation to provide a better means to compare the methods on similar scales. Again, this normalization provides a zero-mean, unit variance distribution.

Using a 95% and 97.5% confidence value for the single and double detection algorithm, respectively, the anomalies are correctly detected at Node01, Node04, Node10 and Node11. The multiple-linear regression method does provide a slightly higher probability of detection on Node01, Node04 and Node10 than the conventional gross count detection algorithm, but both provide a probability of detection greater than 100% for all three nodes due to the high source activity. The anomalous spectra are provided in Figure 69 for each alarm node with an alarm during Event 2. The anomalous spectra at each node is then background-subtracted, filtered, and

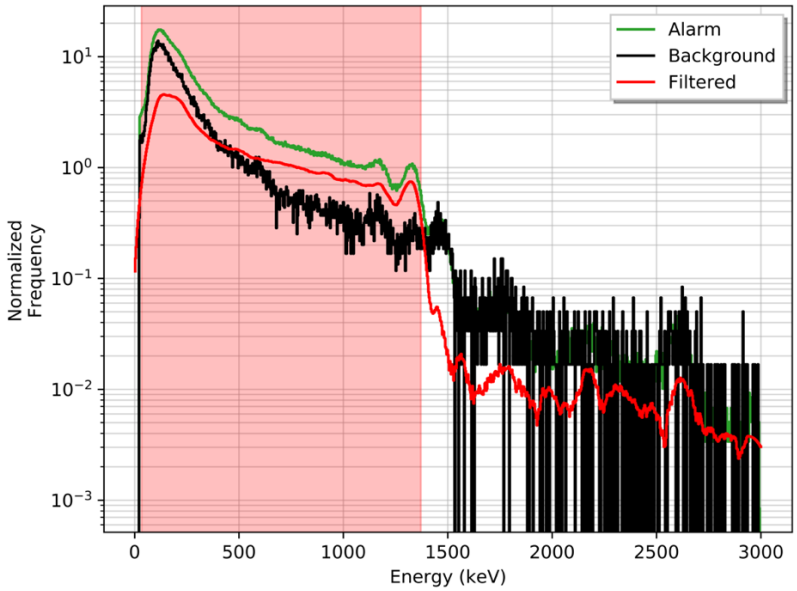


Figure 67: Event 2 region-of-interest from filtered anomalous spectrum at Node06.

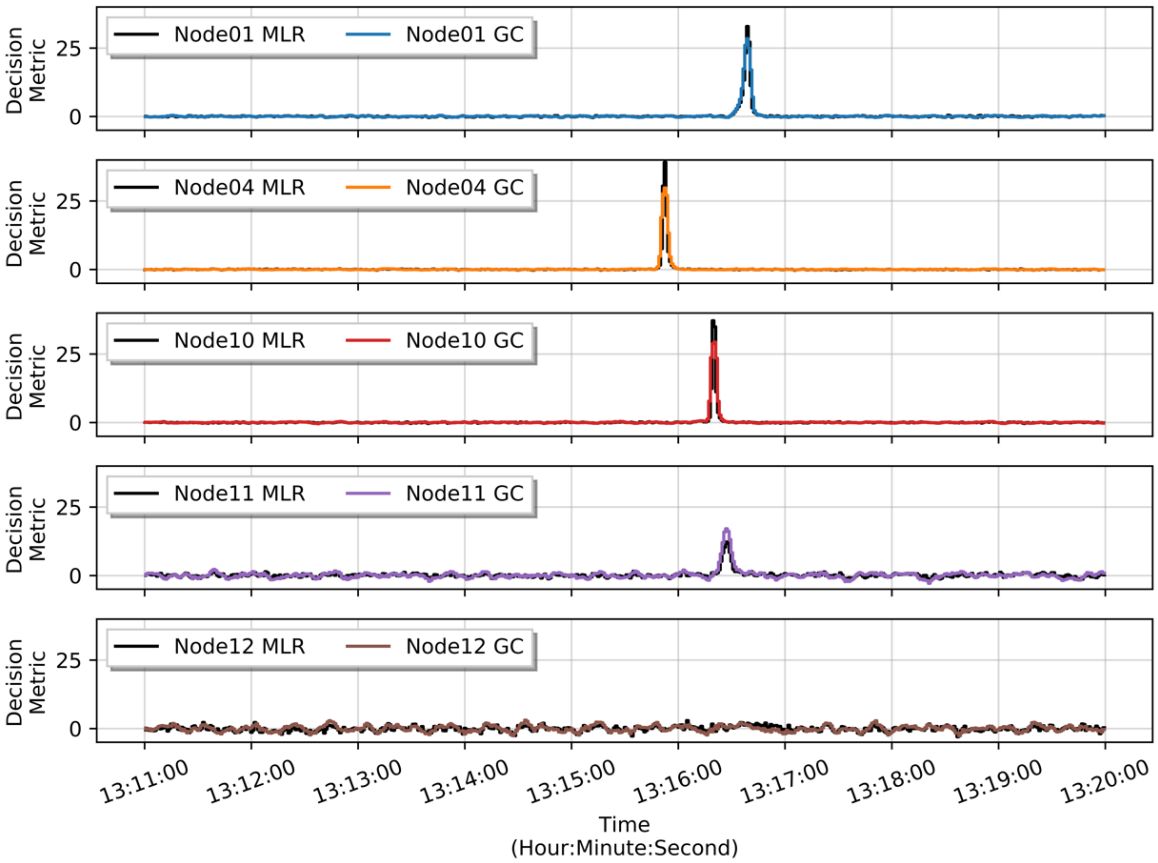


Figure 68: Event 2 decision metric time-series for remaining nodes.

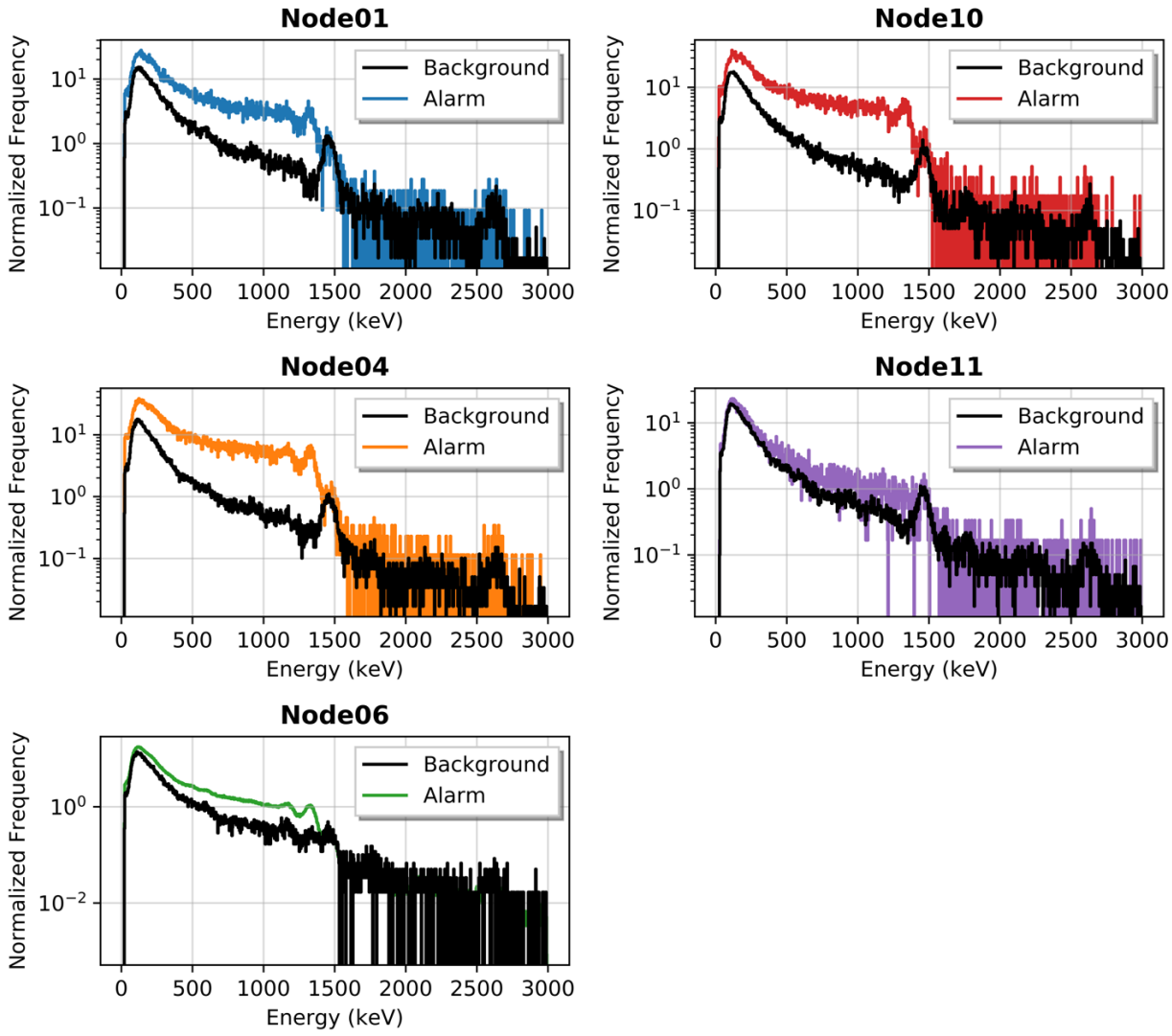


Figure 69: Event 2 alarm and background spectra collected at five nodes.

normalized to provide a more accurate comparison. The resulting spectra are provided in Figure 70.

At lower energies, the peak previously identified around 1.17 MeV, and 1.33 MeV appear to align along with similar intensities in the energy bins below 1.40 MeV. Additionally, a relatively sharp increase at 0.200 MeV is observed which is hypothesized to be a result from shielding around the material in transit as well as a cumulation of overlapping back scatter peaks. Node11 observes large discrepancies between the original anomalous spectrum from Node06 and varies greatly at energies above the region-of-interest. This is contributed to the relative low count rate and a short alarm length observed during the anomalous event at Node11. This results in a sparse energy spectrum with limited spectral information, such as the two identified energy peaks, and large noise at higher energies even after filtering is applied. Again, the intensity is not a factor when employ the cosine similarity method, thus the high normalized frequency in the higher energy region is not of particular concern.

The cosine similarity method is utilized to compare the anomalous spectra collected from Node01, Node04, Node10, and Node11 with the initial anomalous spectra from Node06. The cosine similarity values for each collected spectrum Node01, Node04, Node10, and Node11 to the Node06 spectrum is provided in Figure 71 along with the cosine similarity values for the combined anomalous spectra provided in Table 8. Similar to Event 1, Event 2 observes larger cosine similarity values for the combined anomalous spectra with values slightly greater than 0.990, but still observes relatively high values for background spectra.

Event 3 has a unique alarm sequence to that of Event 1 and Event 2. The radioactive material is initial detected at Node04 then is subsequently detected at Node12 before exiting the area at Node01. The time-series count rate data for the nodes during Event 3 is provided in

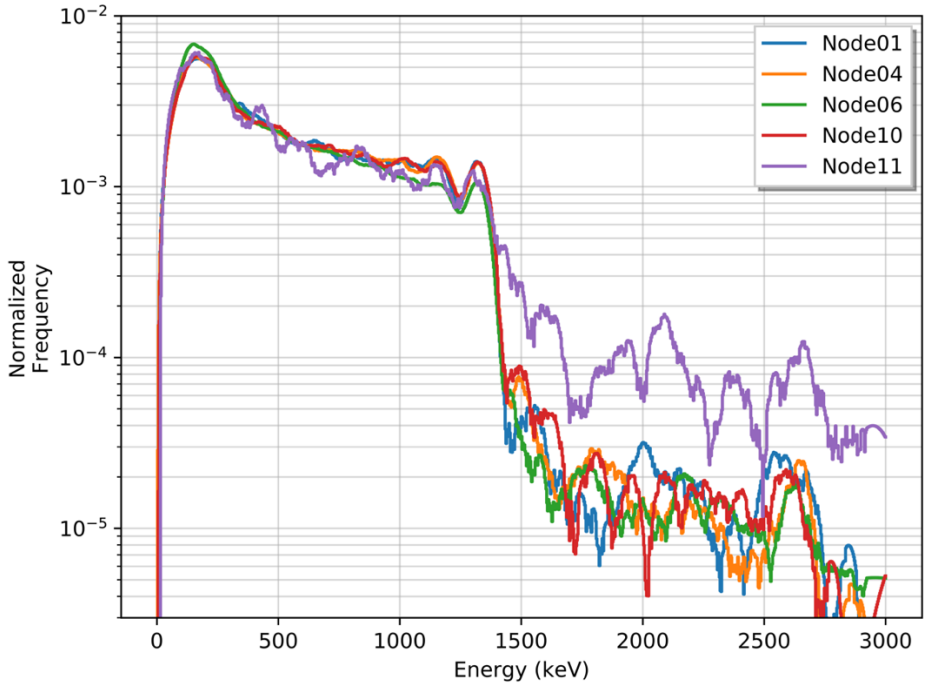


Figure 70: Event 2 anomalous spectra after background-subtracting, filtering, and normalizing.

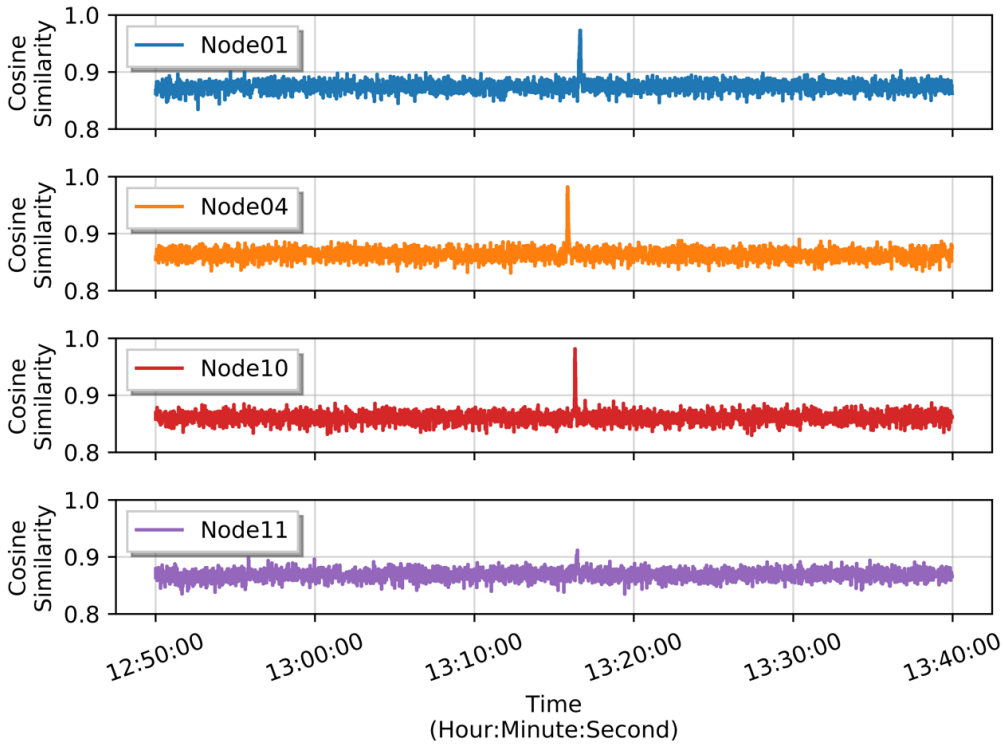


Figure 71: Event 2 cosine similarity time-series.

Table 8: Event 2 cosine similarity values for anomalous spectra.

Node	**Cosine Similarity**
Node01	0.992
Node04	0.994
Node10	0.994
Node11	0.991

Figure 72. The region-of-interest is then calculated from the initial event at Node01. The resulting region-of-interest calculated for Node04, shown as the light-red region in Figure 73, ranges from 0.0 to 1.39 MeV after applying the region-of-interest algorithm to the background-subtracted, filtered anomalous spectrum.

From the initial alarm spectrum at Node04, two energy peaks are observed around 1.17 and 1.33 MeV with an associated down scatter, most likely resulting from ^{60}Co in the radioactive material similar to Event 2. Identical to the previous analysis performed for Event 1 and Event 2, the region-of-interest is utilized in the double detection method where the multiple-linear regression model detection method is operated along with the gross count detection method. The time-series decision metrics for both the single and double detection methods are shown in Figure 74. Again, the decision metrics are normalized to construct a zero-mean, unit variance distribution to compare the decision metric values. Similar to the previous Event 1 and Event 2, with a 95% and 97.5% confidence, both anomalous events are detected with a slight increase in sensitivity in the region-of-interest approach, visible by a higher decision metric.

The anomalous spectra are provided in Figure 75 for each alarm node with an alarm during Event 3. Identical to the previous analysis on Event 1 and Event 2, the anomalous spectra at each node is then background-subtracted, filtered, and normalized to provide a more accurate comparison. The resulting spectra are overlaid together in Figure 76.

At lower energies, the previously identified peaks at around 1.17 and 1.33 MeV appear to align along with similar intensities in the energy bins below 1400 keV. The cosine similarity values for each collected spectrum Node01, Node06, Node10, Node11, and Node12 compared to the Node04 spectrum is provided in Figure 77 along with the cosine similarity values for the combined anomalous spectra from Node01 and Node12 are provided Table 9. Identical to

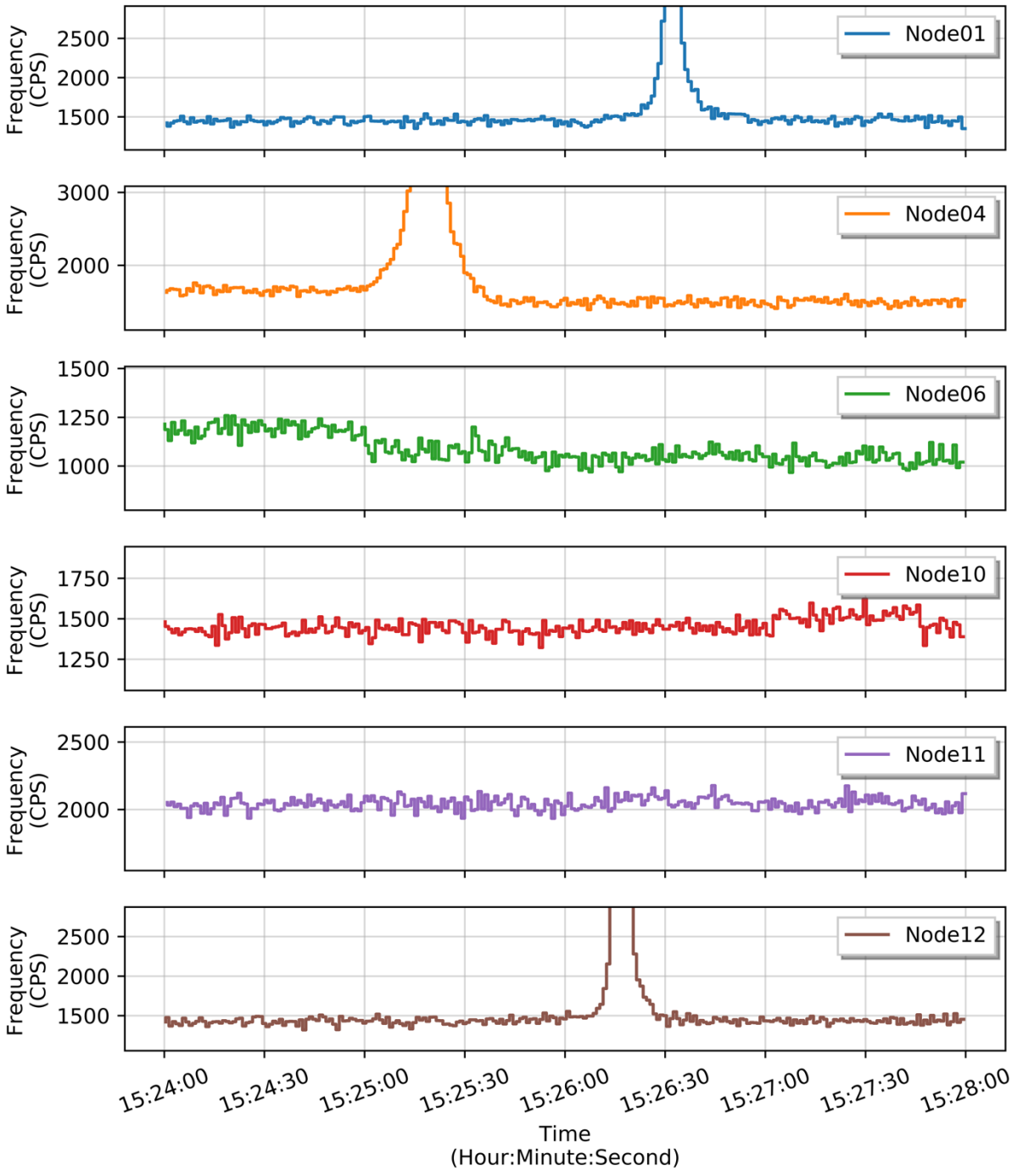


Figure 72: Event 3 time series of count rates observed on the ORNL sensor array.

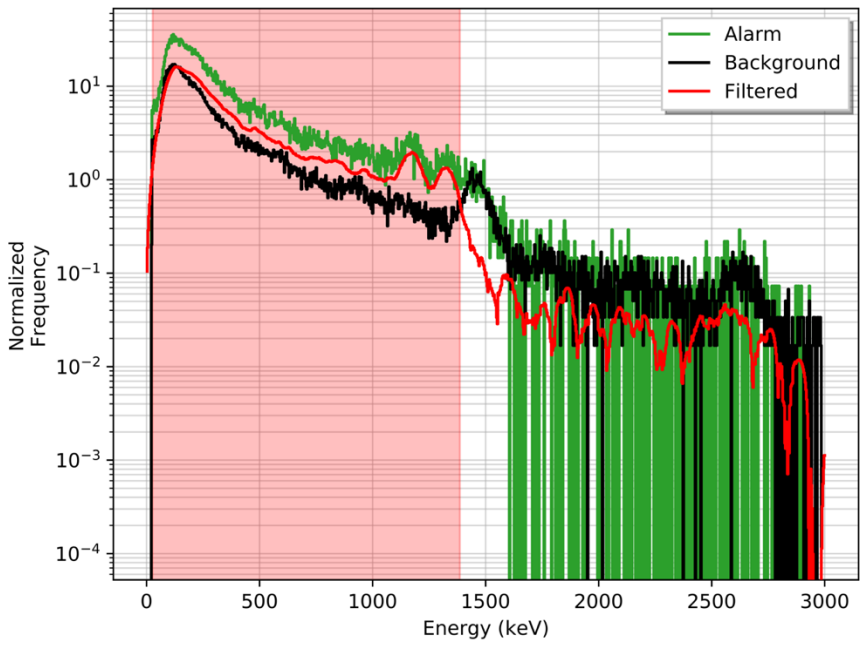


Figure 73: Event 3 region-of-interest from filtered anomalous spectrum at Node04.

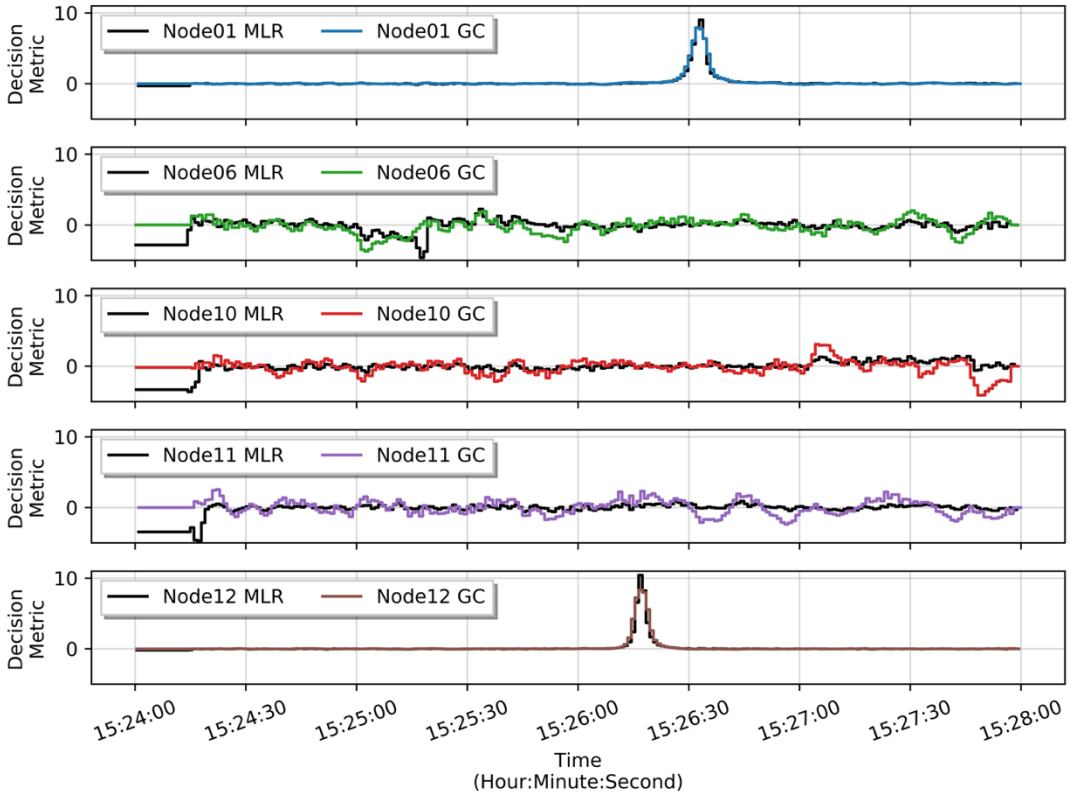


Figure 74: Event 3 decision metric time-series for remaining nodes.

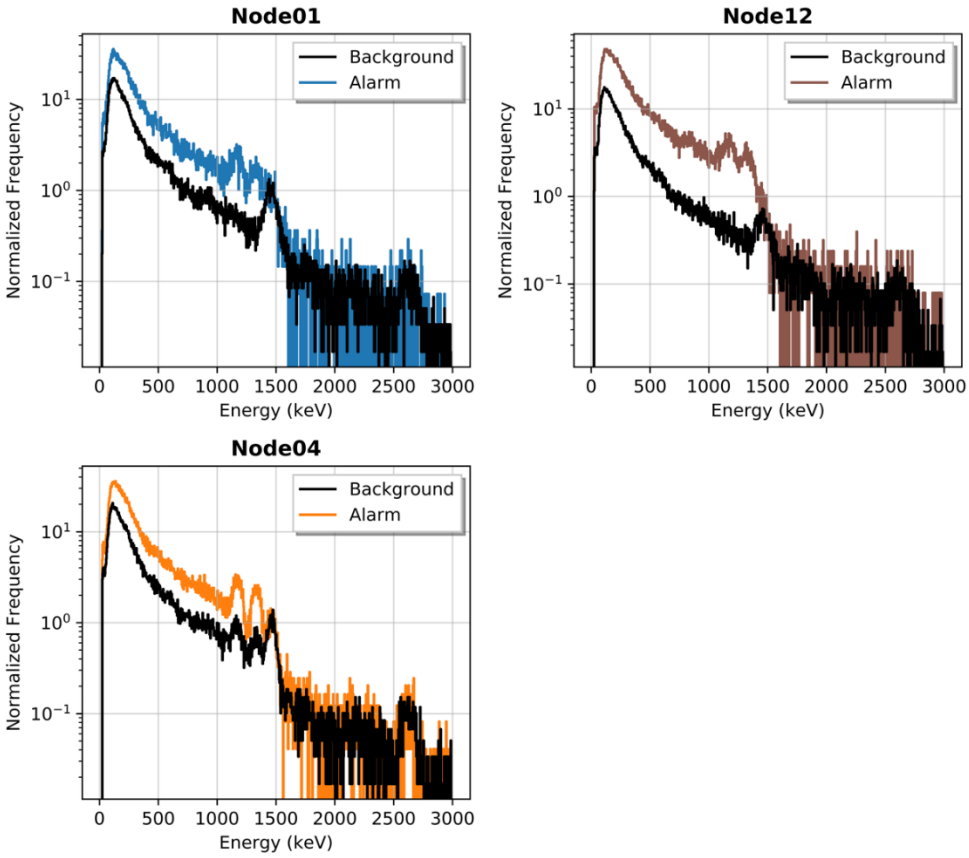


Figure 75: Event 3 alarm and background spectra collected at three nodes.

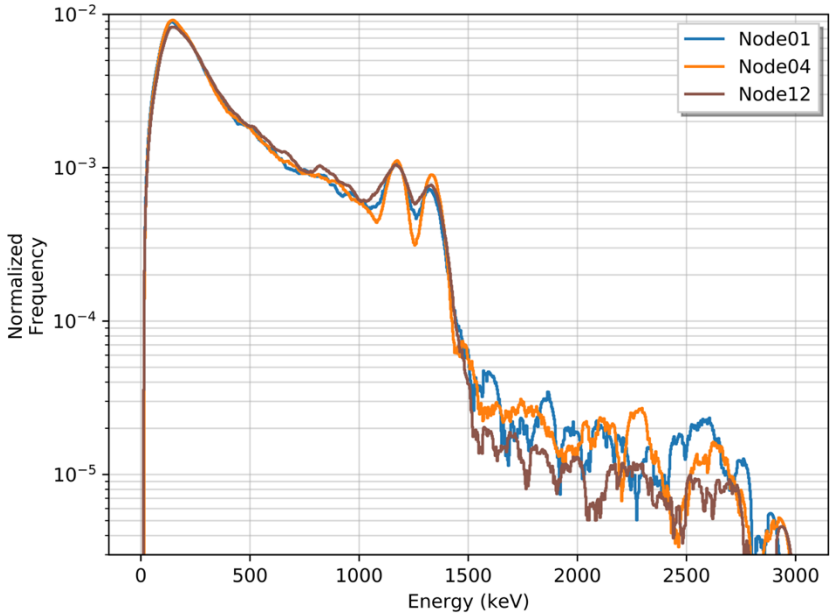


Figure 76: Event 3 anomalous spectra after background-subtracting, filtering, and normalizing.

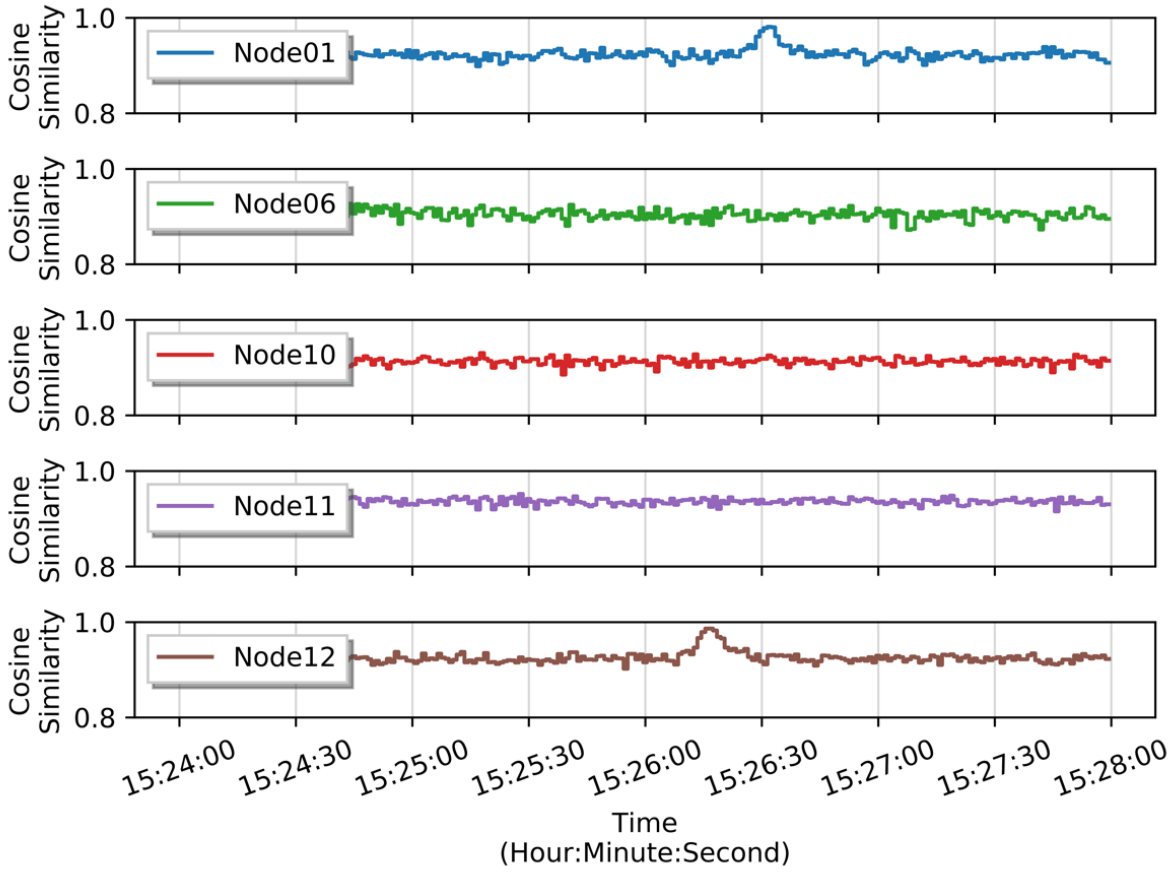


Figure 77: Event 3 cosine similarity time-series.

Table 9: Event 3 cosine similarity values for anomalous spectra.

Node	**Cosine Similarity**
Node01	0.998
Node12	0.997

Event 1 and Event 2, the cosine similarity values in the time-series figure illustrate an interesting issue when attempting to decide whether the collect spectra at each node are samples from the same distribution as the anomalous spectrum from Node04 due to the relatively high cosine similarity values (i.e. greater than 0.85 for nodes in Event 3). Additionally, the combined spectra collected during the anomalous events at Node01 and Node12 provide a much higher cosine similarity values of 0.997 and 0.998, respectively, increasing the confidence that the anomalous spectra are samples from the same radioactive material.

Another extension of this method was observed when utilizing this region-of-interest selection method on real-world data where sources that produced an increase across the entire energy spectrum. This is typically a result of high-energy photon down scatter from neutrons sources interacting with materials between the source and detector. In such cases, a region-of-interest would encompass a majority, or all, of the energy bins in the spectrum, resulting in near-identical detection method in the region-of-interest to the conventional gross counts method. Additionally, the lower count rate typically observed in the higher energy bins can create minute differences in detection sensitivity if the region-of-interest encompasses the majority of energy bins. For this work, a threshold is set to 1.80 MeV, where if the region-of-interest extends beyond the threshold, the region-of-interest methods are not utilized.

4.4: Bayesian Classification Methodology

To remain organized in utilizing raw sensor data and information as it becomes available, this section provides a systematic approach the outputs previously presented on radiation data for robust decision-making. It is important to note that the decision-making process presented are similar to classification processes that are employed for pattern recognition in machine learning with a plethora of literature on numerous unique approaches and background information exists.

Thus, if the reader is interested in such process, the author suggests reviewing both decision theory, pattern recognition, and classification theory [170, 172, 176, 177].

Prior to introducing the decision-making approaches utilized in this work, a few key concepts should be discussed as they create the foundation for the decisions made for the proposed architecture. A significant issue when creating a process to analyze data and make decisions from multiple nodes exists within network communication. Most notably, latency and bandwidth issues can be prevalent and are only further exacerbated when analyzing relatively large datasets at high frequency. Latency quantifies the time delay in transferring data; therefore, a high latency increases the time delay before the data can be transmitted or received. In this work, latency issues are largely ignored as long as the latency either remains relatively small or resolves itself to a lower value in the near feature, before a subsequent event is detected at another node. For distributed sensor networks with large distances between the sensors, the latency should be smaller than the expected travel time between the nodes, which, again, is assumed to not be limiting factor in this scenario. Bandwidth refers to the maximum rate at which data can be transferred and can create a large back log of information that is vital downstream if, for instance, the bandwidth is small relative to the size (e.g., number of bytes) of the data and frequency of transfer. This matter can further be exacerbated if the bandwidth begins to shrink as new nodes are operated simultaneously on the same network. Bandwidth issues can be minimized, if not virtually alleviated, by reducing the size of the data to be transferred over the network. The data queued for transfer over the network can be decreased by processing the data collected from the sensors locally at each node. This type of distributed computing paradigm for minimizing the bandwidth and detection response time locally is often referred to as edge computing [178, 179]. The processed data products, or features, can

subsequently be transferred rather than the raw data (e.g., full energy spectra), therefore reducing the total size of the data transferred. Furthermore, unlike in decades past, the current state of computational and processing power in commercially available electronics (e.g., single-board computers) provides the ability to process multiple data streams on a single system. This work assumes all sensor data can be efficiently processed locally, requiring no cloud computing or central processing center.

Thus far, the sequential methodology for analyzing radiation data and tracking anomalies by means of increasing detection sensitivity via the region-of-interest approach and comparing the anomalous spectra via the cosine similarity can be generally explained by Figure 78.

Prior to beginning an anomalous event sequence, the data from the primary sensor, in this case a radiation detector, is initially analyzed using the conventional gross counts detection methodology. Upon detection of an anomalous signal using a predetermined decision threshold for that system, the two previously discussed features are extracted: (1) region-of-interest that optimizes the signal-to-noise and (2) background-subtracted, filter anomalous spectra. That information is to be appropriately packaged (i.e. compressed and/or encrypted) and transferred to other nodes on the network. Subsequently, the other nodes in the network utilize the information to improve detection sensitivity and classify if an observed anomaly is similar to the initial anomaly, similar to the process provided in Figure 79. Making a classification by comparing data can be accomplished by performing a series of hypotheses tests on three propositions:

H_0 : no anomaly is observed;

H_1 : the local anomaly observed is the same as the initial anomaly for the event; and

H_2 : the local anomaly observed is different than the initial anomaly for the event.

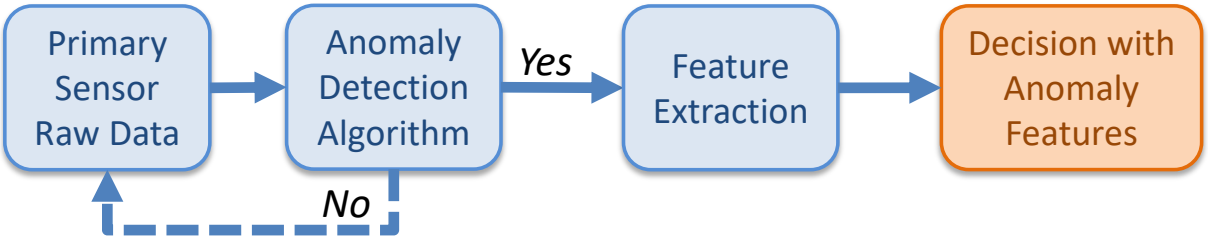


Figure 78: Schematic of single node process for processing radiation sensor data.

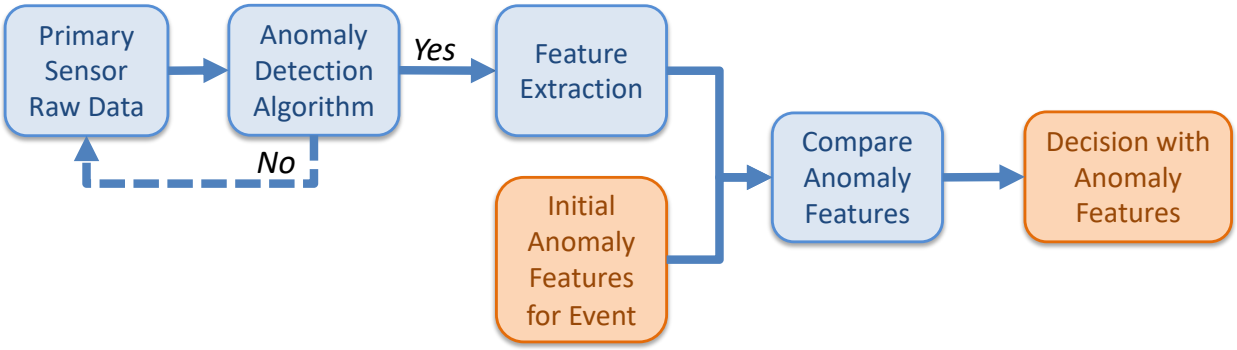


Figure 79: Schematic of single node process for processing radiation sensor data utilizing previous information.

In such scenarios, the specified propositions, or hypotheses, should be Boolean with either True or False values selected by from a numerical degree of belief. Constructing and testing hypotheses is utilized various reasoning processes and can be extended to update the system beliefs when new information becomes available for statistical inference. This can be accomplished by constructing a Bayesian relationship between the evidence (i.e. output from sensors) and the hypotheses. Bayesian methods for decision theory, often referred to as Bayes classifiers, provide a paradigm to systematically deal with proposition uncertainties when making decision. Bayesian analysis utilizes Bayes' theorem, provided in the following equation, which is constructed by taking advantage of definitions of the conditional and joint probabilities for the given evidence (E) and hypotheses (H), $P(\boldsymbol{H}|\boldsymbol{E}) = \frac{P(\boldsymbol{E}|\boldsymbol{H})\,P(\boldsymbol{H})}{P(\boldsymbol{E})}$. A full derivation is provided in the Appendix.

In many scenarios, such as fusing data from multiple sensors, the evidence and hypotheses parameters are written in vector form. For instance, for the evidence vector $\boldsymbol{E} \in R^d$ (i.e. feature vector in d-dimensional space) and j-number of hypotheses, the original equation is written as $P(H_j|\boldsymbol{E}) = \frac{P(\boldsymbol{E}|H_j)\,P(H_j)}{P(\boldsymbol{E})}$. This process allows for the probability that a hypothesis (H) is true given the evidence (E) to be numerically calculated with values typically between 0 and 1, if the normalization probability $P(E)$ is used to scale the numerator. The normalization probability in the dominator is calculated through the summation of (j) possibilities if the probabilities are discrete, as show in the following equation. In contrast, if the probabilities are continuous probabilities, where a the data can take any value in a region of the distribution, the summation changes to an integral of the lower and upper bounds of the region $P(H_j|\boldsymbol{E}) = \frac{P(\boldsymbol{E}|H_j)\,P(H_j)}{\Sigma_j P(\boldsymbol{E}|H_j)\,P(H_j)}$.

The higher the posterior probability for a hypothesis, the more likely that the respective hypothesis is supported by the evidence. Thus, a common method for selecting a hypothesis is to select the hypothesis with the highest posterior probability, which is referred to as the *maximum a posteriori*, or MAP, decision rule when selecting the mode of the posterior distribution. The denominator can be removed to simplify the calculation as users are often times interested in the hypothesis prediction rather than the actual likelihood value. For instance, the following decision rule could be employed for a two-hypothesis test regime.

Decision 1: Choose H_0 if $P(\boldsymbol{E}|H_0)\ P(H_0) > P(\boldsymbol{E}|H_1)\ P(H_1)$

Decision 2: Otherwise, choose H_1 for $P(\boldsymbol{E}|H_0)\ P(H_0) \leq P(\boldsymbol{E}|H_1)\ P(H_1)$

Or, in a more shorthand approach with multiple hypotheses, this stated as the following.

Decision: $H^* = \text{argmax}_\text{j}\ P(H_j|\boldsymbol{E})$

This is a natural approach to govern the decision space by selecting the decision solely based on the posterior probability, $P(H_j|\boldsymbol{E})$. Worth noting, this method will observe distinct solutions in various scenarios. For instance, for a given observation $\boldsymbol{E}$, the likelihood functions for two hypotheses are equal (i.e. $P(H_0|\boldsymbol{E}) = P(H_1|\boldsymbol{E})$), this decision-making approach is governed by the priors $P(H_0)$ and $P(H_1)$. Additionally, if the priors for two hypotheses are equal, this approach is governed solely by the likelihoods.

Using the information previously discussed, classification and decision-making frameworks can be thought of as a network that performs an action based on a function's output of the data, or evidence, vector. Such frameworks are relatively common for pattern classifiers utilizing both parametric and nonparametric methods. In this context, these functions are typically referred to as discriminant functions and perform an operation that assigns a score to the data from the input space. The discriminant functions $g(\boldsymbol{X})$ can create decision boundaries or

surfaces that partition the data input space governed by the vector $\boldsymbol{X}$. The primary goal of these functions is to obtain an optimal decision based on the data in the input space [170]. These frameworks are often shown as block diagram or network that illustrates the general process flow from a d-dimensional input vector and c discriminant functions to decision or classification; example diagram provided in Figure 80. Similar to the aforementioned MAP methodology, the highest value output from the discriminant function is used to selects the decision as described by the following; choose decision $\boldsymbol{D}_i$ if $g_i(\boldsymbol{X}) > g_j(\boldsymbol{X})$ for all $j \neq i$. In the pure Bayesian context, the discriminant function is generally selected to minimize the risk of taking an action (i.e. minimize the probability of taking a wrong action). In other words, this process produces a classification decision based on estimates of the state of the system represented as outputs of the discriminant functions. This Bayesian risk construct is created by introducing a conditional loss function $\lambda(\alpha_i|H_j)$ if, in this case, the true state of the system is H_j and action a_i is chosen. Given $P(H_j|\boldsymbol{X})$, which represents the probability that the true state of the system is H_j given the data vector $\boldsymbol{X}$, and conditional loss function, a conditional probability of the risk of choosing action α_i given the data vector $\boldsymbol{X}$ is represented as $R(\alpha_i|\boldsymbol{X}) = \sum_{\boldsymbol{j=1}}^{\boldsymbol{c}} \lambda(\alpha_i|H_j)\, P(H_j|\boldsymbol{X})$ [170].

The overall risk is minimized by selecting action α_i that provides the smallest conditional risk $R(\alpha_i|\boldsymbol{X})$, which is commonly referred to as Bayes risk [170]. A vital and sometimes difficult step is selecting the appropriate loss function. Primarily for its simplicity, a zero-one loss function can be employed to minimize the error rate by multiplying the conditional probability $P(\omega_j|\boldsymbol{X})$ by either 0 or 1 for an incorrect or correct decision, respectively, representing all errors are equally costly. This is expressed as the following piecewise function, $\lambda(\alpha_i|H_j) =$

$$\begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \text{ for } i, j = 1, \dots, c \text{ .}$$

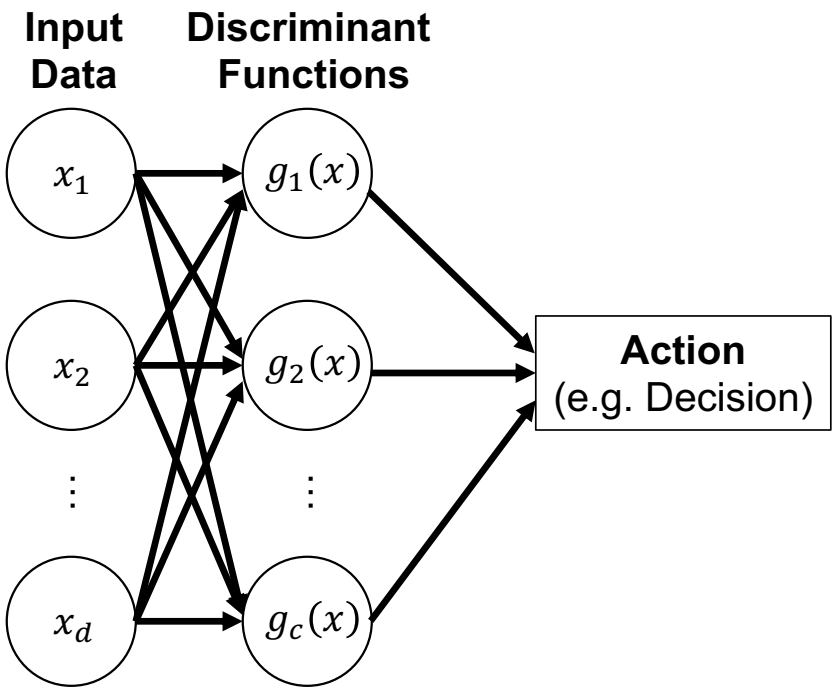


Figure 80: Example network diagram for decision-making process.

Inserting the zero-one loss function into the previous equation for the conditional risk, a simplification is observed due to all incorrect actions being eliminated $R(\alpha_i|\boldsymbol{X}) = \sum_{\boldsymbol{j=1}}^{\boldsymbol{c}} \lambda(\alpha_i|H_j)\, P(H_j|\boldsymbol{X}) = \sum_{\boldsymbol{j\neq i}}^{\boldsymbol{c}} P(H_j|\boldsymbol{X}) = 1 - P(H_j|\boldsymbol{X})$**.** With this relationship, it is observed at the conditional risk is minimized when the posterior probability $P(\omega_j|\boldsymbol{X})$ is maximized. Thus, using the zero-one loss function results in an identical decision rule as the previous MAP decision rule. Using this relationship, the minimum conditional risk or the posterior probability can be utilized as the discriminant function to minimize the conditional risk or the error rate, respectively (i.e., $g_i(x) = -R(\alpha_i|\boldsymbol{X})$ and $\mathrm{g_i}(x) = P(H_j|\boldsymbol{X})$). The error rate minimization is often presented as the product of the likelihood and prior probability, which recall is equal to the posterior probability after removing the evidence probability $P(\boldsymbol{X})$ that normalizes, or scales, the product. Furthermore, the natural logarithm is commonly utilized on the discriminant function resulting in a summation rather than a product as shown in $\ln g_i(x) = \ln\left(P(\boldsymbol{X}|H_j)\, P(H_j)\right)$ and $\ln g_i(x) = \ln P(\boldsymbol{X}|H_j) + \ln P(H_j)$.

Prior to creating the decision surfaces for the data collected from the radiation sensors, it is vital to first construct the likelihood and prior probabilities. Continuing with the framework previously provided in Figure 80, the input data vector $\boldsymbol{X}$ contains the features from the radiation data stream; namely, the cosine similarity. In this work, two additional features are added to the input layer that represent estimates of the probability of the count rate is an anomaly and the time of arrival probability to the sensor. The notations for each are parameter in the input data vector $\boldsymbol{X}$ are described in the following list.

x_0 : cosine similarity value between initial anomaly spectrum and local spectrum;

x_1 : probability observed count rate is an anomaly; and,

x_2 : expected time-of-arrival initial anomaly.

With the parameters now defined, the original Bayesian equation can be expanded to reveal how the actual posterior $P(H_j|\boldsymbol{X})$ is calculated, as shown in the following. $P\left(H_j\middle|\boldsymbol{X}\right) = \frac{P\left(\boldsymbol{X}\middle|H_j\right)P(H_j)}{P(\boldsymbol{X})} = \frac{P\left(\boldsymbol{X}\middle|H_j\right)P(H_j)}{\sum_i P(\boldsymbol{X}|H_i)\,P(H_i)}$ and, when expanding the data vector $\boldsymbol{X}$, $P\left(H_j\middle|\, x_0, x_1, x_2\right) = \frac{P\left(x_0\middle|H_j\right)P\left(x_1\middle|H_j\right)P\left(x_2\middle|H_j\right)\,P(H_j)}{\sum_i P(x_0|H_i)\,P(x_1|H_i)\,P(x_2|H_i)\,P(H_i)}$. Again, this assume the likelihood functions for the three parameters are independent of one another, which eliminates the conditional probabilities between the data parameters. Of note, assuming the data parameters are independent in this way is often referred to as the naïve Bayes method. If solely the output values of the posterior are utilized in that the maximum value will be selected as the classification, the normalization factor in the denominator will no longer be required to scale the values between 0 and 1, thus, $P\left(H_j\middle|\, x_0, x_1, x_2\right) = P\left(x_0\middle|H_j\right)P\left(x_1\middle|H_j\right)P\left(x_2\middle|H_j\right)P\left(H_j\right)$.

However, in this work, the posterior probabilities will be sequentially updated, thus requiring the normalizing factor in the denominator to scale the values. At this point, the conditional probabilities for each data parameter and the prior distributions for the three hypotheses need to be calculated. The prior distribution will be sequentially updated by utilizing the previous time step's posterior but requires an initial value. The initial values are selected rather heuristically by using the following values, where $x_{1,B}$ represents the background, or historic, count rate data $P(H_0) = P\left(x_{1,B}\middle|H_0\right)$, $P(H_1) = 0.01,$ and $P(H_2) = P\left(x_{1,B}\middle|H_2\right)$. Although an original assumption in this work is that no signal will be simultaneously observed at multiple node, the probability that the current state is *same anomaly*, or $P(H_1)$, is specified as 0.01, rather than zero. This eliminate a zero probability from existing after obtaining the product in the numerator of Bayes' formula. The $P(H_0)$ and $P(H_2)$ prior distributions are selected to reflect the system state is governed initially by the observed count rate for *no anomaly* and

different anomaly, respectively. This method for initializing $P(H_0)$ and $P(H_2)$ provides a straightforward method for calculating the specific values by using a previous time step's background count rate data, denoted at $x_{1,B}$, rather than arbitrarily selecting a fixed value. Lastly, the prior probabilities are renormalized to force the $P(H_1)$ to the specified value of 0.01. The conditional probabilities will be constructed in the following sections for each parameter in the data vector $\boldsymbol{X}$.

4.4.1: Cosine Similarity Conditional Probability

Recall that the cosine similarity values attempt to provide a degree of belief that the local spectrum is a sample from the same radioactive material observed in the initial anomaly. It is shown in the cosine similarity time-series figures (i.e. see Figure 71, Figure 74, and Figure 77) that normal background spectra values remain relatively high with a mean values around 0.85 for all three examples. The actual anomalous spectra observed similarity values greater than 0.97 for the individual, 1-second anomalous spectra and similarity values greater than 0.99 for the combined anomalous spectra in all three events. With this information, the conditional probabilities can be constructed. A normal, or Gaussian, distribution can be utilized to construct the conditional probability, which requires two parameters to be selected; namely, mean and variance. Each conditional probability is addressed in order for $P(x_0|H_0)$, $P(x_0|H_1)$, and $P(x_0|H_2)$.

First, constructing the conditional probability that the cosine similarity value is observed given that system state is *no anomaly* (H_0) is created by utilizing by the cosine similarity values prior to the initial anomaly being detected. This allows for the expected cosine similarity mean and variance for H_0 to be estimated solely based on each sensor's background spectrum and produces unique results for each sensor. This step is vital as the background energy spectra can

drastically vary spatially. For all three examples, 30-seconds of spectral data is utilized to obtain an estimate of the mean and variance. In this particular case, all three examples had cosine similarity near identical means and variances with expected values at approximately 0.85 and 0.009 for the mean and variance, respectively.

Second, the conditional probability that the cosine similarity value is observed given that system state is *same anomaly* (H_1) is created by coupling the observed values from the anomalies in the three examples with the theoretical expectation. Recall that if the two samples under comparison using the cosine similarity approach, a perfect match equates to a value of 1.0. Thus, the expected mean of the conditional probability is set to 1.0 with a variance 0.0025, which allows for values to slightly differ from an exact match and limits the values around the typically observed *no anomaly* values. This conditional probability is a truncated normal distribution as values above 1.0 are not possible, resulting in only the lower half of the distribution being considered.

Lastly, the conditional probability that the cosine similarity value is observed given that system state is *different anomaly* (H_2) requires more detail prior to constructing. With no prior information regarding detecting a radioactive material with unknown spectral features, the resulting conditional probability must remain robust to encompass all possible scenarios, even a different radioactive material with a similar signature as the initial. Furthermore, the cosine similarity methodology has yet to be utilized for a different radioactive source in this work. Recall that for radioactive materials with spectral features in the higher energy region of the spectrum (e.g., >1500 keV) and no features in the lower energy region will generally maintain similarity values as the photon downscatter region will remain similar in shape. Two scenarios from the ORNL sensor array, referred to as *Different Alarm* in this scenario, are provided to

examine the impact different spectral feature in the lower energy region impact the cosine similarity values. Both are compared to other anomalies detected on Node04 from Event 3 (see Figure 73 and Figure 75), which is represented as *Initial Alarm* in this scenario. The first example (Figure 81) is a radioactive material with a prominent photopeak around 60 keV most likely resulting from a radioactive material with Am-241 (i.e. gamma-ray emission at 59.5 keV). The second example (Figure 82) is the observed Node04 anomalous response during Event 1 shown previously (see Figure 61) that shows relatively large deviations from the initial alarm from 500 keV to 1100 keV.

The cosine similarity values between the different anomalous spectra Scenario 1 and Scenario 2 and the initial alarm are 0.747 and 0.987, respectively. These results highlight the reasoning for using a conditional probability for H_2 to encompass a multitude of scenarios, including larger discrepancies, such as Scenario 1, and similar, or even nearly identical, sensor responses from a different radioactive material, such as Scenario 2. To cover the regions in the probability space, a normal distribution with a relatively large variance is utilized with a mean between the expected values for the *no anomaly* and *same anomaly* hypotheses. For this work, the mean and variance are obtained for this hypothesis by selected a mean 5% higher than the background, or *no anomaly*, values and a variance equal to 1% of the selected mean. These values are selected as they encompass the physical expectations of this hypothesis, but more optimal solutions may be possible if a full sensitivity analysis were to be performed. Additionally, a different radioactive material with identical, or near-identical, photon emissions could enter the area being monitored by the sensor array, which could force the Bayesian framework to classify the different material as *same anomaly* based on the high cosine similarity

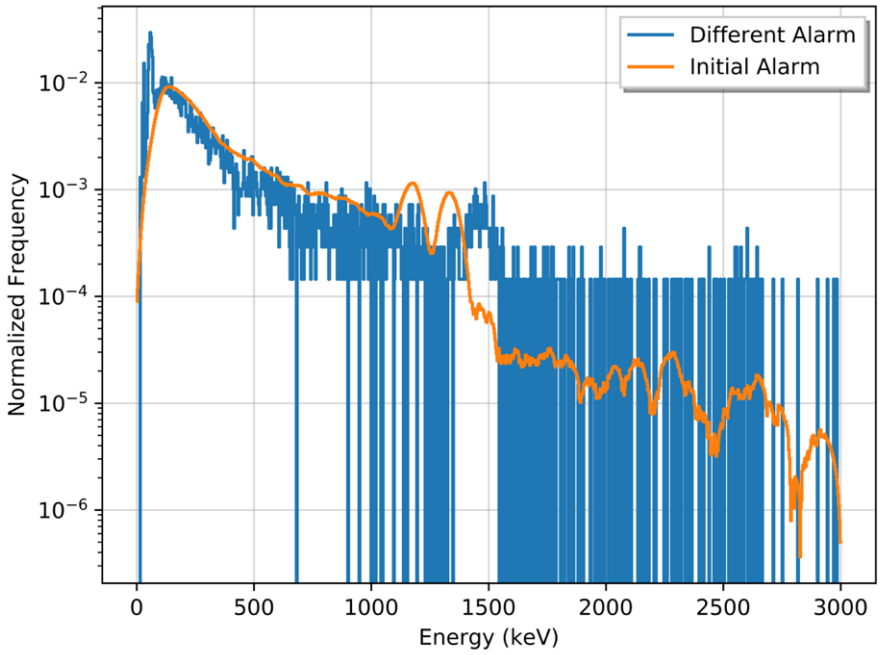


Figure 81: Energy spectra for Scenario 1 comparing different anomalous spectra from Node04.

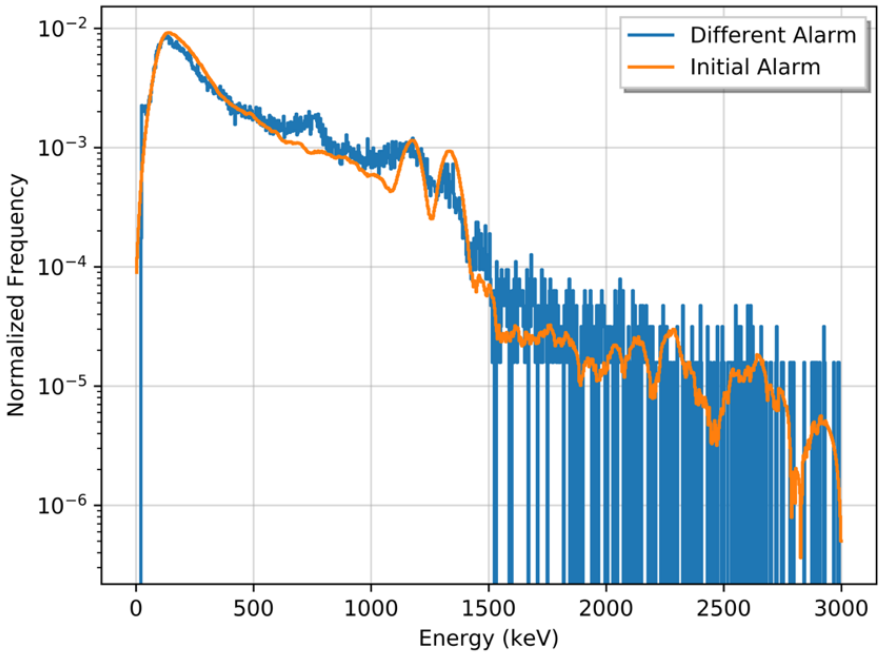


Figure 82: Energy spectra for Scenario 2 comparing Event 1 and Event 3 anomalies from Node04.

values. Such a scenario is deemed highly unlikely by the author and is not addressed in this work, but it is worth noting.

With these methods for calculating the means and variances of the conditional probabilities for each hypothesis, the probabilities can be constructed assuming a normal distribution. For instance, this approach applied to Node01 in Event 1 provided the following mean and variance for each hypothesis. The tilde symbol (~) means that the parameter, or in this instance the conditional probability, *is distributed as* the function provided, N is a normal distribution, and E represents the expected value.

$$P(x_0|H_0) \sim P\left(x_0 \middle| N(\mu_0, \sigma_0^2)\right) \text{ with } E(\mu_0), E(\sigma_0^2) = 0.85, 0.010$$

$$P(x_0|H_1) \sim P\left(x_0 \middle| N(\mu_1, \sigma_1^2)\right) \text{ with } E(\mu_1), E(\sigma_1^2) = 1.00, 0.003$$

$$P(x_0|H_2) \sim P\left(x_0 \middle| N(\mu_2, \sigma_2^2)\right) \text{ with } E(\mu_2), E(\sigma_2^2) = 0.91, 0.010$$

Using these parameters, the conditional probabilities are provided in Figure 83. The probabilities do not exist above a cosine similarity value of 1.0 as a value greater than 1.0 is impossible and, more importantly, Figure 83 shows the amount of overlap between the probabilities. Recall that a large overlap between distribution decreases a classifier's sensitivity and potentially increasing the false-positive rate. Utilizing other parameters can alleviate this issue or, at the very least, increase the system's sensitivity.

4.4.2: Count Rate Conditional Probability

Given that the radiation emission process follows a Poisson process, the probability that an observed count rate is an outlier can be relatively trivial for static scenarios. As a discrete distribution, the Poisson distribution can be utilized to calculate expected time to obtain a specific number of events given the mean of the rate (μ) for the data $P(k) = e^{-\mu}\frac{\mu^k}{k!}$, representing the probability of observing *k* events in the time period. Using this relationship, the rolling-window

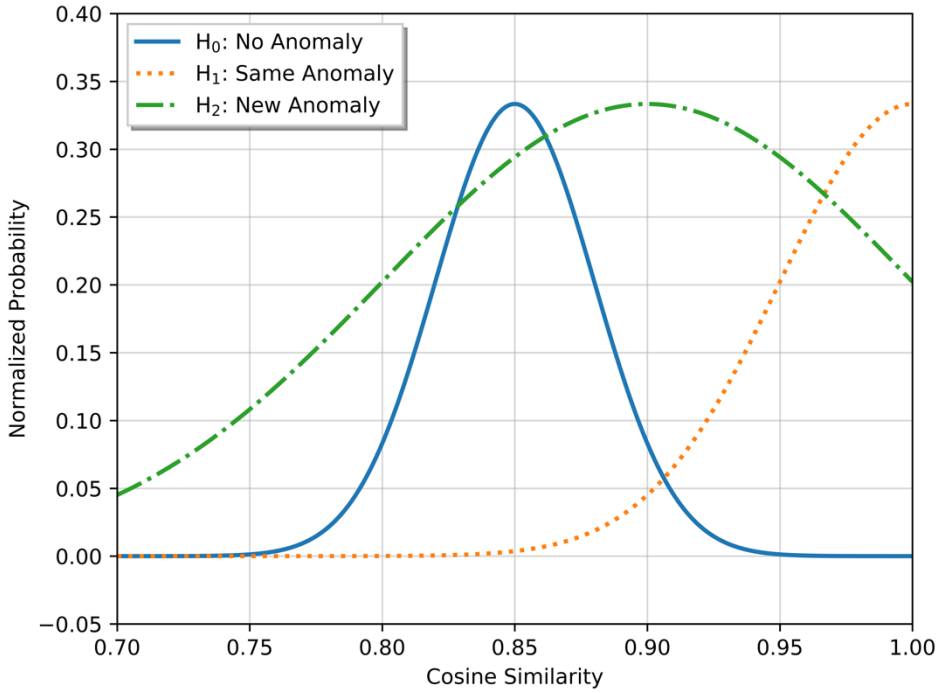


Figure 83: Cosine similarity conditional probability for Node01 in Event 1.

method is utilized to obtain the probability that the observed count rate is an anomaly from the expected, background count rate. For instance, assume the mean count rate (μ_{CR}) in the background window is 1800 and a value of 1875 is observed. The probability of observing a value of ≤1875 is 96%, which means the probability of observing a value ≥1875 is 4%, as shown in Figure 84. In the context of this work, count rate anomalies are expected to be above the mean of the background as this work does not address count rate suppression anomalies. Therefore, the conditional probabilities for $P(x_1|H_1)$ and $P(x_1|H_2)$ can be expressed as opposite of $P(x_1|H_0)$ above its mean. Both of the conditional probabilities $P(x_1|H_1)$ and $P(x_1|H_2)$ are equal in this regime, as no information is utilized to discriminant between the two hypotheses solely based on the observed count rates. Additionally, in the absence of any contextual information regarding the relative position of the radioactive material to the radiation sensor, no estimate for the activity can be made and, thus, the likelihood values remain equal for all large count rate values. Using the previous example for a mean count rate of 1800, Figure 85 shows the conditional probabilities normalized to sum to 1.0.

The conditional probabilities are only shown above the mean as count rates below the mean are not classified as anomalies. Furthermore, in order to address an issue that will arise in later sections of this work, an increase of 0.01 to the conditional probabilities $P(x_1|H_1)$ and $P(x_1|H_2)$ is applied in order to eliminate the probabilities from being exactly zero. This addition provides a minute change to the distribution by allowing the posterior probabilities to exist for all real values without greatly impacting the overall probabilities. As a result, the conditional probability $P(x_1|H_0)$ below the mean can then be expressed in $P(x_1|H_0) = 0.98$ for $x_1 \leq \mu_{CR}$.

It worth noting that modelling radiation count rate as a Poisson random variable may introduce significant error in predicting anomalies as other parameters other than the radiation

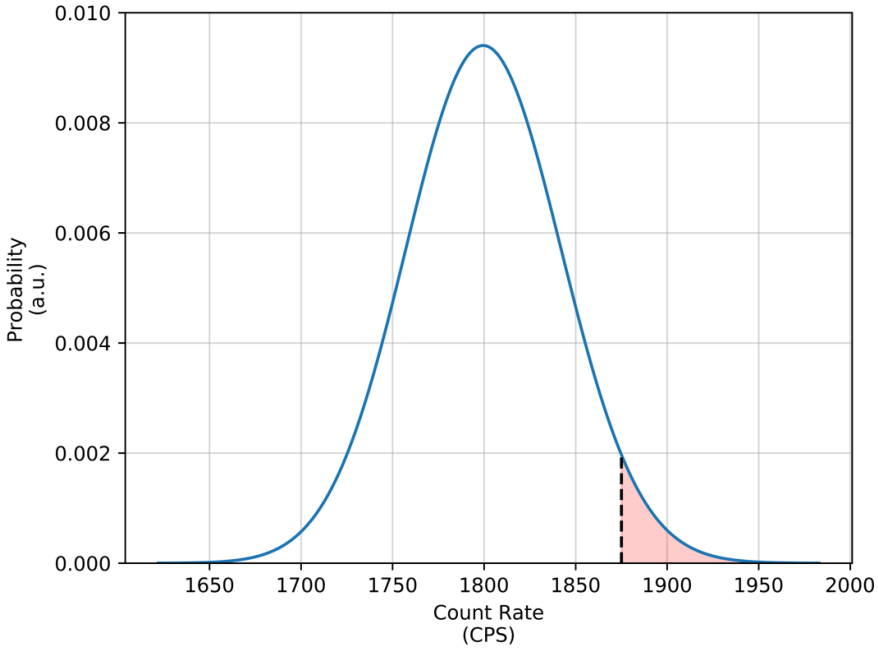


Figure 84: Example Poisson distribution for arbitrary count rate data.

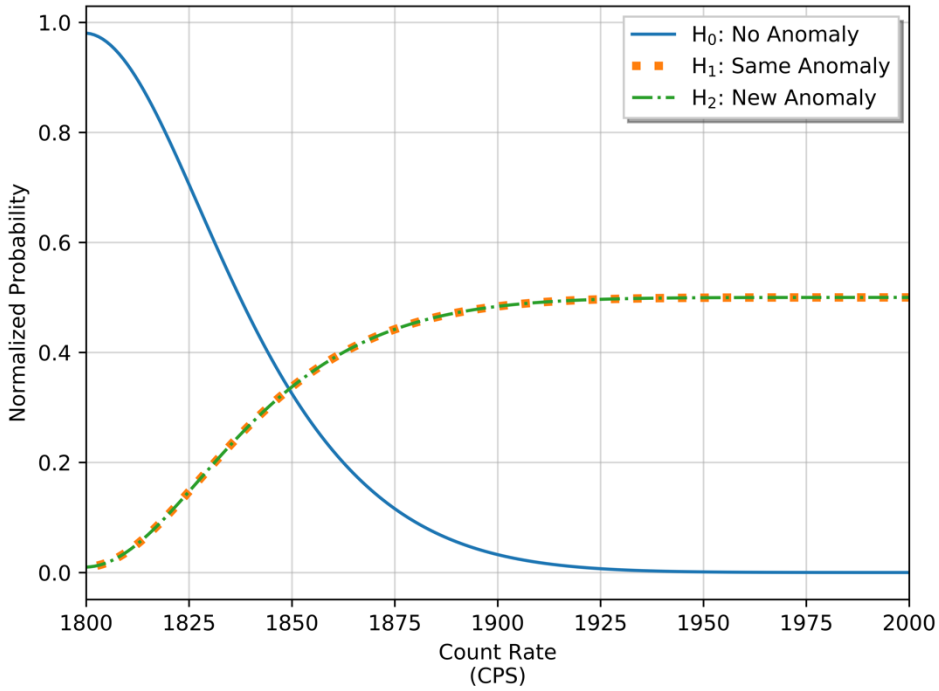


Figure 85: Count rate conditional probabilities for arbitrary count rate data.

emission process contribute to the observed count rate, such as fluctuations in system's electronics, dynamic environment in the vicinity of the system, or weather patterns. These fluctuations increase the variance in the data, widening the distribution. To overcome these issues as normal distribution could be utilized as a substitute for the Poisson as the mean count rate is typically large enough and, along with the variance, can easily be obtained from historic data in a rolling-window. As a continuous distribution, the normal distribution is not able to calculate an exact probability of observing a discrete sample.

4.4.3: Expected Time-of-Arrival Conditional Probability

In the early stages of this work, an observation was made when analyzing data from distributed sensors in an attempt to correlate observed signals across the network in the absence of any contextual data or ground truth information. Nonphysical observations were found in which a radioactive material (e.g., medical radioisotope) was detected from a sensor in the network and subsequently detected at a distant location in a relatively short time period. The expected travel time between the two sensors was far greater than the elapsed time between the two similar signals (e.g., 2 seconds between sensor alarms positioned 10 miles apart along a roadway). To overcome this issue initially, a minimum elapsed time after the initially anomaly was manually specified. Rather than having to manually specify a minimum elapsed time, the conditional probabilities can be constructed to remove nonphysical scenarios. Unfortunately, this concept only effects the $H_1: same\ anomaly$, as the other two hypotheses are equally likely, or uniform, as time elapses.

Assuming that the expected travel time between sensors on the network is known via onboard GPS modules or the node positions are known beforehand, the conditional probability $P(x_2|H_1)$ can be constructed using an exponentially modified normal distribution. This function

is created using the left, lower portion of a normal distribution and applies an exponential decay after the peak at the mean value. The variance used for the normal distribution should reflect the expected travel time, primarily effecting the probability at lower time values. The exponential tail is necessary in this case as expected travel times can vary drastically between two points due to a multitude of reasons (e.g., traffic, weather, stops along the route, etc.) and multiple pathways can exist, all of which result in a longer-than expected travel time.

To add context, assume an expected travel time (μ_t) between the sensor that initially observed the anomaly and a subsequent sensor in the network is 15 minutes. Additionally, it is assumed that the expected travel time is obtained from a reputable source (e.g., live time traffic maps) increasing the confidence in that expected travel time, thus a standard deviation (σ_t) of 20% of the mean (i.e. 3 minutes, in this instance) is utilized. A decay constant for the exponential tail is obtained by assuming the probability is 25% of the mean at twice the mean; calculated as via $T = T_0 e^{-\lambda t}$ and $\lambda(0.25) = -\frac{\ln(T/T_0)}{2\,t_{\exp}} = -\frac{\ln(0.25)}{2*15\text{min}}$. A piecewise function is created to combine the two function over the specific sub-domains of x_2. $P(x_2|H_1) =$

$$\begin{cases} N(\mu_t, \sigma_t^2) & \text{if} \quad t \leq \mu_t \\ T_0 \exp(-\lambda\, t) & \text{if} \quad \mathrm{t} > \mu_t \end{cases}$$

. The conditional probabilities $P(x_2|H_0)$ and $P(x_2|H_2)$ are modelled as uniform distributions as the likelihood of observing the two scenarios are equally likely at any time. To sum to a likelihood of 1.0, the $P(x_2|H_0)$ and $P(x_2|H_2)$ conditional probabilities are calculated by $1 = P(x_2|H_0) + 2\big(P_{1,2}\big)$ where, $P_{1,2} = P(x_2|H_1) = P(x_2|H_2)$ resulting in $P_{1,2} = \frac{1-P(x_2|H_0)}{2} = P(x_2|H_1) = P(x_2|H_2)$.

Using this information along with the assumptions made for this example scenario, the conditional probability $P(x_2|H_1)$ is constructed along with the uniform distributions for

$P(x_2|H_0)$ and $P(x_2|H_2)$, as shown in Figure 86 after normalization of the conditional probabilities.

Worth mentioning, the $P(x_2|H_2)$ conditional probability could be modified if the expected likelihood for a observing a new anomaly is significantly low. For instance, if the sensor network is positioned in a remote area that typically observes a low anomaly rate due to the locations being monitored, the conditional probability $P(x_2|H_2)$ could be decreased, as the likelihood is not equal to $P(x_0|H_2)$. This would increase $P(x_2|H_1)$ while decreasing $P(x_2|H_0)$ and thus, would subsequently impact the posterior probabilities $P(H_j|\boldsymbol{X})$ calculated. For a proof-of-concept, this work utilizes the previous methods of assuming equal, uniform distributions for $P(x_2|H_0)$ and $P(x_2|H_2)$ and does not address altering the $P(x_2|H_2)$ conditional probability as it is operational dependent.

4.4.5: Overview of Bayesian Framework Procedure

This section compiles the previous sections to provide a brief, more organized procedure from the initial anomaly event to extracting the data features from the radiation data, constructing the conditional probabilities, and calculating the posterior probability using the Bayesian framework. Prior to detecting sensor data as an anomaly, each node in the network follows the procedure shown in Figure 87.

In an effort to limit the communications and data traffic over the network, the methods described in the previous sections for determining the region-of-interest to increase detection sensitivity along with calculating a combined background-subtracted spectrum during the observed anomaly, the resulting timestamped information can be compressed and transmitted to other nodes in the network. The subsequent nodes in the network use the features from the initial anomaly to construct the conditional probabilities and perform the hypothesis test to classify the

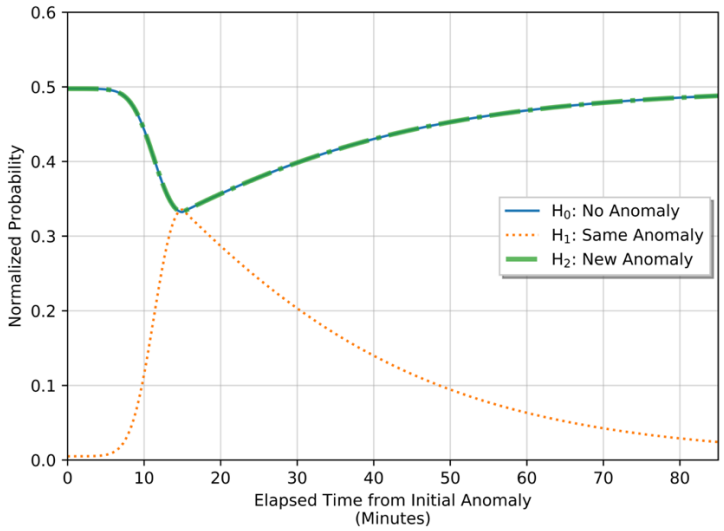


Figure 86: Expected time-of-arrival conditional probabilities for arbitrary data.

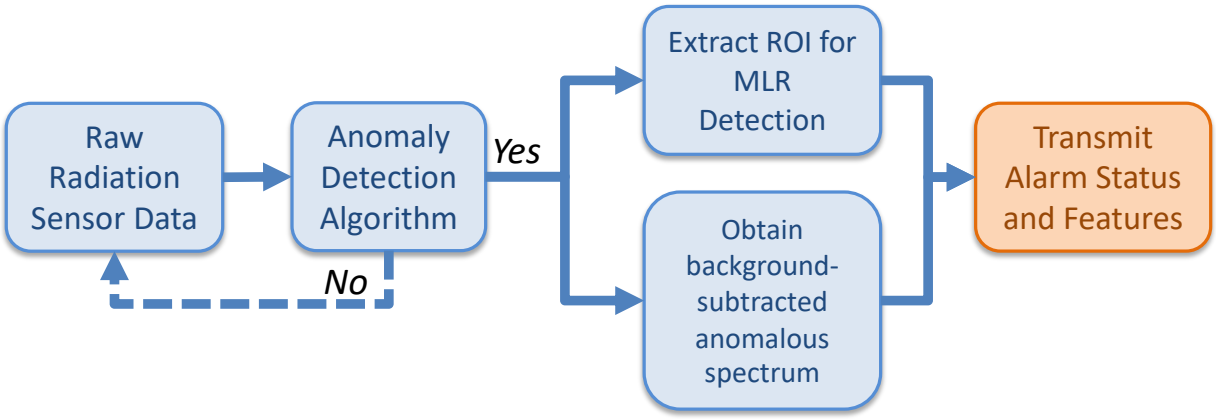


Figure 87: Schematic of initial alarm node procedure.

local node's information as one of the three hypotheses. Figure 88 provides an overview of this process. The exact procedure utilized to construct the initial prior and conditional probabilities is expressed in the following Figure 89. This procedure does require normalization in order scale the probabilities to sum to unity and is performed in later analysis of the previous three example events from the ORNL sensor array. The posterior $P(\boldsymbol{X}|\boldsymbol{H})_t$ of the current timestamp (t) is used as the following timestamp $(t+1)$ prior $P(\boldsymbol{H})_{t+1}$ to continuously update the system belief. Additional context is provided in the following section where the processes are utilized on the three example scenarios previously described from the ORNL sensor array.

4.4.6: Bayesian Framework on ORNL Sensor Array

Using the methods described in the previous sections to construct the conditional probabilities, the posterior probability, which recall is the optimal discriminant function when minimizing the error rate (see page 147), is calculated for the three anomalous events on the ORNL sensor array presented in the previous section. The examples are use the identical designation of Event 1, Event 2, and Event 3 and analyzed in the following pages. For reasons that will be addressed in shortly, the analysis performed on Event 3 is present first followed by Event 1 and Event 2.

Recall that the alarm sequence for Event 3 started at Node04 and subsequently observed at Node12 and Node01. Using the process outlined in Figure 89, the radiation data is utilized to construct the prior and conditional probabilities. Approximations for the time-of-arrival are created by assuming the radioactive material is travelling at the local speed limit of 20 MPH for a linear distance calculated from the GPS modules at each node. Upon initialization of the probabilities, the posterior probabilities for each of the three hypotheses are calculated and sequentially updated as new data is obtained. The resulting time-series, normalized posterior probabilities for Event 3 are provided in Figure 90.

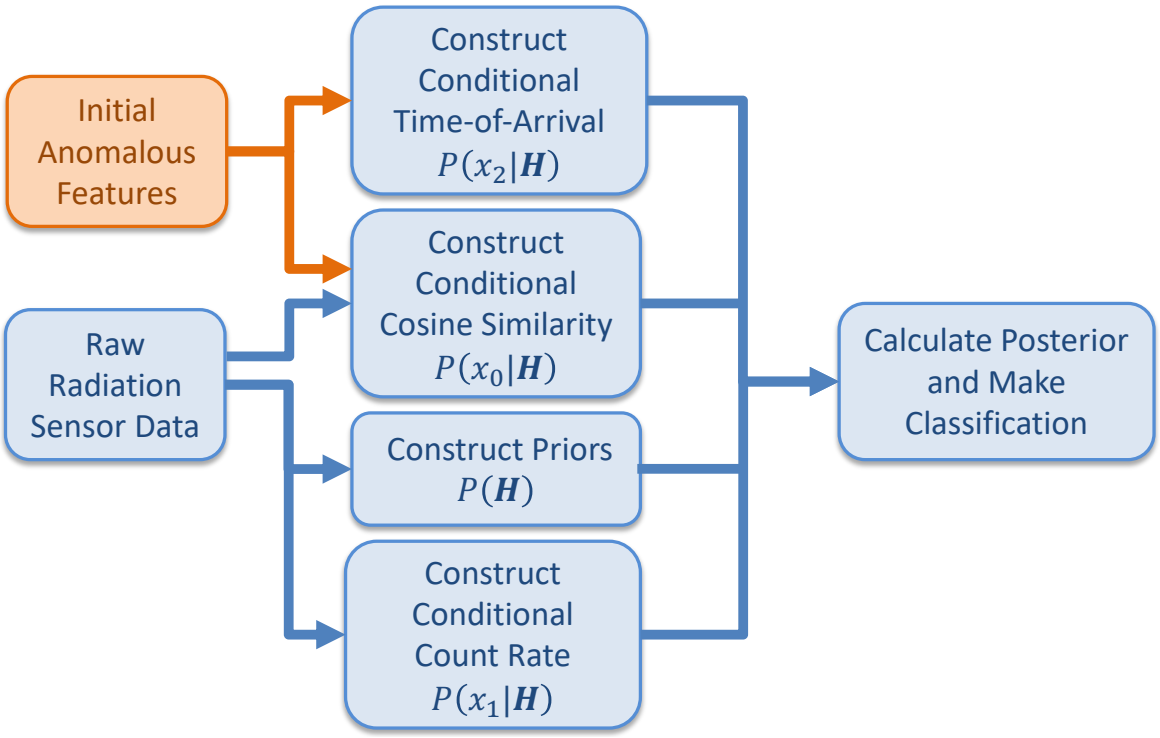


Figure 88: Schematic for constructing Bayesian beliefs after initial alarm.

Initialize Probabilities for Bayesian Framework

- *Cosine Similarity Data* : $P(x_0|\boldsymbol{H})$
  - $P(x_0|H_0)$: Use last 30-seconds to create normal distribution; $N(\mu_{CS}, \sigma_{CS}^2)|_0^1$.
  - $P(x_0|H_1)$: Construct left-side normal distribution; $N\left(1.0, \sqrt{0.0025}\right)|_0^1$.
  - $P(x_0|H_2)$: Construct normal distribution; $N\left((\mu_{CS} + 0.05\mu_{CS}),\ 0.1(\mu_{CS} + 0.05(\mu_{CS}))\right)|_0^1$.
- *Count Rate Data* : $P(x_1|\boldsymbol{H})$
  - $P(x_1|H_0)$: Use last 30-seconds of data to obtain mean count rate; $Poisson(\mu_{CR})$.
  - $P(x_1|H_1)$: $0.5\left(1 - P(x_1|H_0)\right)$
  - $P(x_1|H_2)$: $0.5\left(1 - P(x_1|H_0)\right)$
- *Time-Elapsed Data* : $P(x_2|\boldsymbol{H})$
  - $P(x_2|H_0)$: $0.5\left(1 - P(x_2|H_1)\right)$
  - $P(x_2|H_1)$: Obtain approximate arrival time and create piecewise function.
  - $P(x_2|H_2)$: $0.5\left(1 - P(x_2|H_1)\right)$
- *Prior Distributions*: $P(\boldsymbol{H})$
  - $P(H_0)$: $P(x_1|H_0)$
  - $P(H_1)$: 0.01
  - $P(H_2)$: $P(x_1|H_2)$

Figure 89: Initialization process of prior and conditional probabilities for Bayesian framework.

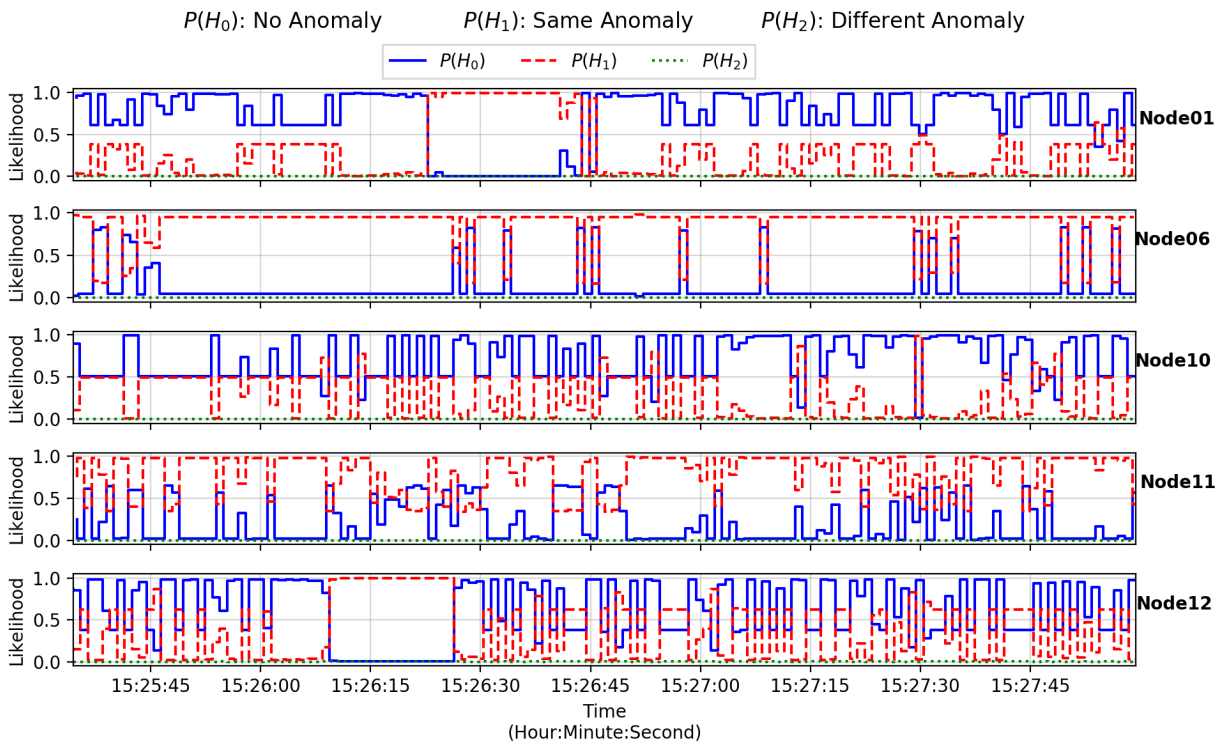


Figure 90: Event 3 posterior probabilities using the Bayesian framework.

The observed posterior probabilities alternate between $P(H_0)$ and $P(H_1)$, even at locations where the radioactive material was not detected (e.g., Node06, Node10, and Node11). This is primarily attributed to the high cosine similarity values between the normal background energy spectra and the initial anomalous spectra from Node04 (see Figure 77 for cosine similarity time-series data). The average cosine similarity values for the nodes ranged from 0.91 to 0.94 for the normal background energy spectra. To overcome this issue, the conditional probabilities are weighted to better represent their importance in the hypothesis test. If the cosine similarity values for the historic 30-second normal background energy spectra are above 0.90, the prior distributions are weighted more heavily towards the observed count rate and lowered for both the cosine similarity and the time-of-arrival. The specific weights are provided in the following relation for a given mean cosine similarity value (μ_{Cos}) for the background energy spectra.

If $\mu_{Cos} \geq 0.90$: $P(x_0|\boldsymbol{H}) = 0.10\big(P(x_0|\boldsymbol{H})\big)$, $P(x_1|\boldsymbol{H}) = 0.80\big(P(x_1|\boldsymbol{H})\big)$, $P(x_2|\boldsymbol{H}) = 0.10\big(P(x_2|\boldsymbol{H})\big)$; Else: $P(x_0|\boldsymbol{H}) = 0.40\big(P(x_0|\boldsymbol{H})\big)$, $P(x_1|\boldsymbol{H}) = 0.40\big(P(x_1|\boldsymbol{H})\big)$, $P(x_2|\boldsymbol{H}) = 0.20\big(P(x_2|\boldsymbol{H})\big)$.

As shown in the previous relation, the weights are also applied during scenarios with lower cosine similarity values. Additionally, the weighting scheme lowered the dependency on the time-of-arrival data, as these values can vary drastically vary and is not expected to have a high contribution to the classification process. This weighting scheme proposed will be utilized moving forward. The results for Event 3 using the updated weighting scheme are shown in Figure 91. After applying the proposed weighting scheme, the anomalies observed at Node12 and Node01 are correctly identified and classified as H_1 *same anomaly* at 15:26:10 and 15:26:28, respectively. Additionally, no false positives are observed on any of the nodes, increasing the confidence that the proposed weight scheme improves the classification performance.

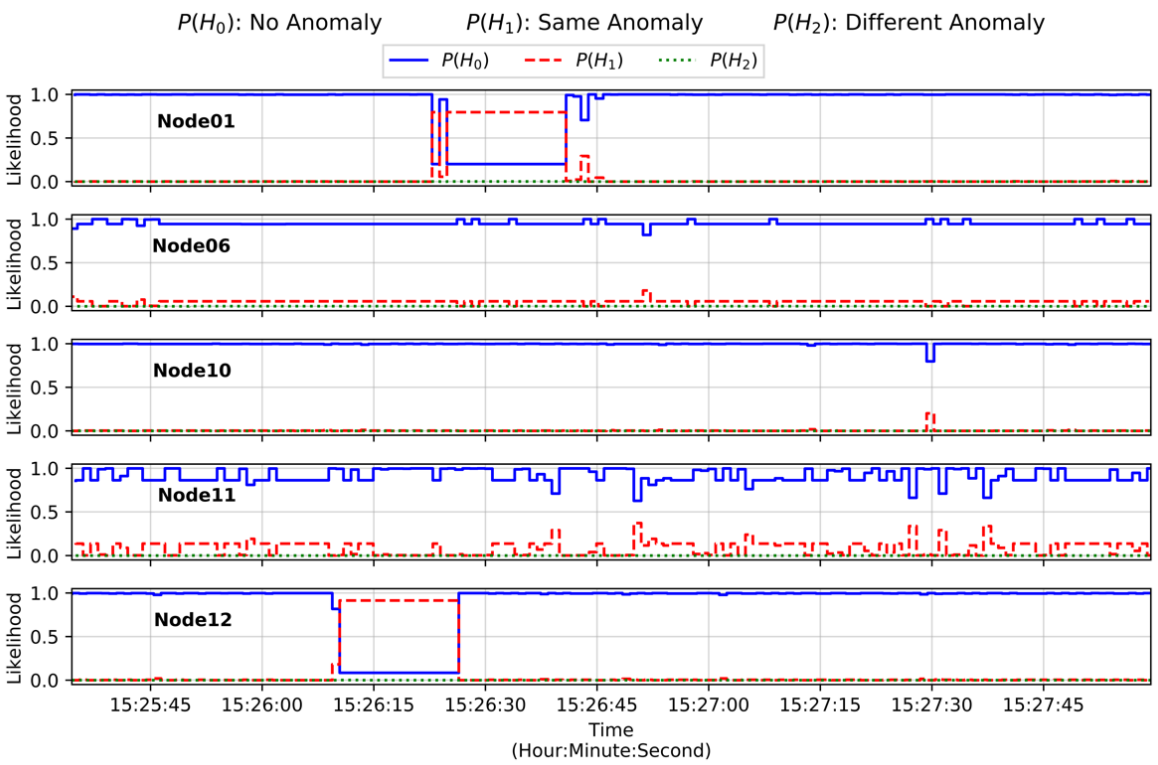


Figure 91: Event 3 posterior probabilities using the Bayesian framework with the weighted probability approach.

Switching back to Event 1, the initial anomaly was observed on Node06, thus the remaining nodes in the array are analyzed using the Bayesian framework. An identical process to the previous analysis for Event 3 is utilized to construct the prior and conditional probabilities. However, in contrast to Event 3, the cosine similarity values between normal background energy spectra and Node06's initial alarm spectra all lie below 0.90 resulting in the different weighting scheme shown previously. The resulting time-series, normalized posterior probabilities for Event 1 are provided in Figure 92.

Recall that the true alarm sequence for Event 1 begins at Node06 and progresses to Node04, Node10, and Node01, in that order. Node04 and Node01 correctly identify two anomalies shown by the $P(H_1)$ *same anomaly* hypothesis being dominant at approximately 13:09:50 and 13:10:38, respectively. More interesting is the misclassification occurs at the beginning and end of the anomalies at Node10 and Node0, respectively, where the $P(H_2)$ *different anomaly* hypothesis is dominant. This is primarily attributed to the moderate cosine similarity values for normal background at Node10 (i.e. 0.87 compared to 0.82 and 0.84 for Node01 and Node04, respectively) resulting in less sensitivity in discriminating between the $P(H_1)$ and $P(H_2)$ due to the increased overlapping of their respective conditional probabilities, $P(x_0|H_1)$ and $P(x_0|H_2)$. Furthermore, at lower count rate anomalies, less spectral feature will be observed which will generally result in a slight change in the cosine value until higher count rates are observed providing more features. No solution is proposed to overcome this issue, which the author believes highlights the need for additional data features, say features extracted from other sensor modalities, to alleviate misclassifications and false positives.

In Event 2, the radioactive material was initially detected at Node06 before travelling to Node04, Node10, Node11, and Node01, in that specific order. Similar to Event 1, the cosine

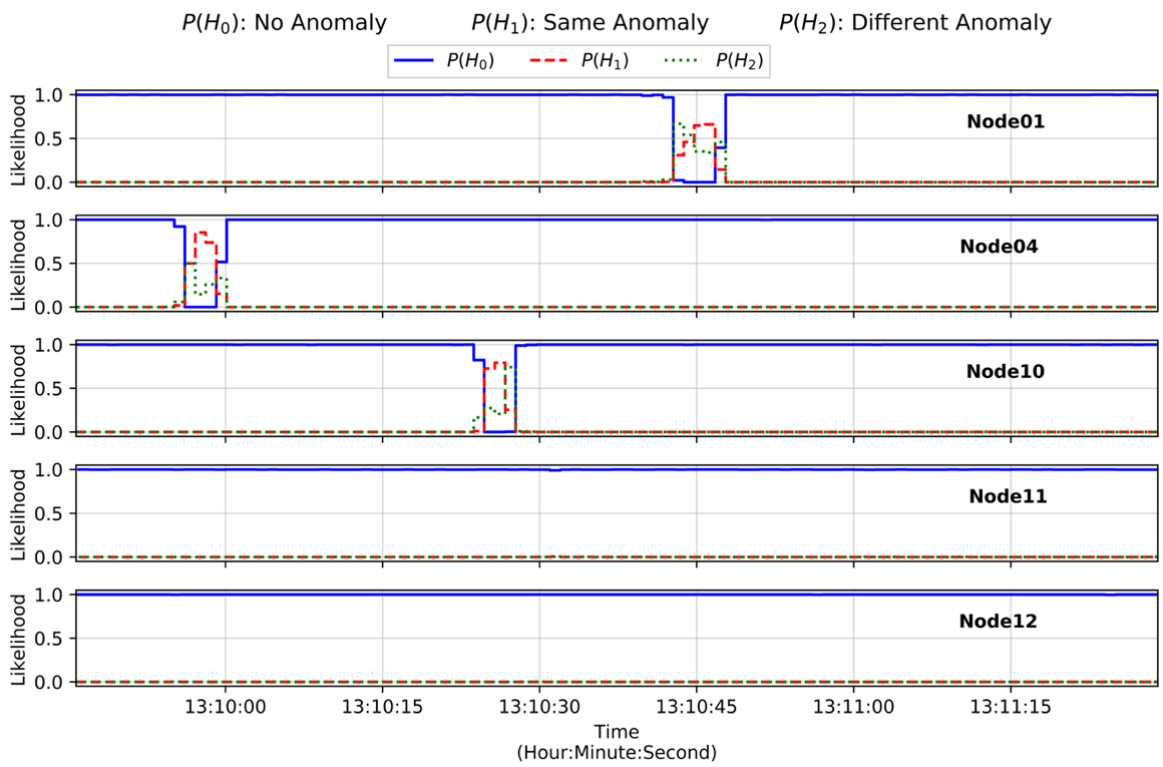


Figure 92: Event 1 posterior probabilities using the Bayesian framework.

similarity values between the normal background spectra and anomalous spectra observed from Node06 lie well below 0.90 (see Figure 71) . The time-series, normalized posterior probabilities for Event 2 are provided in Figure 93. All the alarms were identified with proper $P(H_1)$ *same anomaly* classifications at Node04, Node10, Node11, and Node01 with no false positives. The $P(H_2)$ *different anomaly* probabilities do observe an increase during the anomalous events, but is not the dominant, or highest, classification during the anomalies.Similar to the Event 1, the initial spectra observed during the anomalous events provides a cosine similarity value between *no anomaly* values $P(x_0|H_0)$ and *same anomaly* values $P(x_0|H_1)$, which is within the *different anomaly* $P(x_0|H_2)$ dominant region.

A key takeaway from this analysis is highlighted by the competition between the *same anomaly* and *different anomaly* during an anomalous event containing the same, or very similar, radioactive material. This is primarily due to the approaches utilized when constructing the conditional probabilities for the three data features used in this framework. This issue is inherent to scenarios that solely utilize radiation sensors and lack contextual information, forcing the use of uninformative or overlapping probabilities to that of the *same anomaly* probabilities, $P(x_0, x_1, x_2|H_1)$. Thus, the system's sensitivity in discriminating between the two hypotheses decreases.

4.4.7: Bayesian Framework Multiple Source Injection Study

While the current methodology using the Bayesian framework with the updated weighting scheme provides a more robust approach than the unweighted scheme, a sensitivity analysis should be performed to further examine how the framework operates in other scenarios. For instance, in the three events previously discussed, each anomaly is known to be observed at subsequent nodes originates from the same radioactive material as the initial anomaly. No

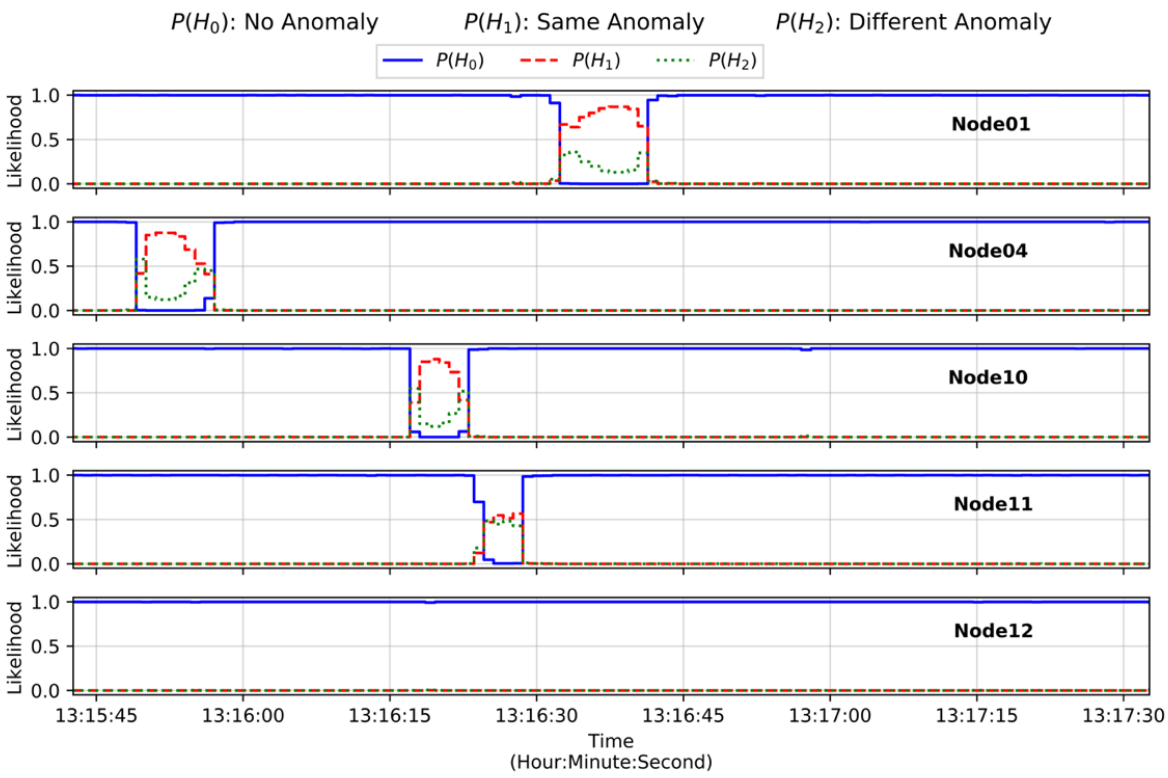


Figure 93: Event 2 posterior probabilities using the Bayesian framework.

anomaly sequence yet has analyzed the impact of an additional radioactive material entering the area after the initial anomaly is detected. This is addressed by inserting energy spectra from a different radioactive material into the data stream for Event 2. Two different source spectra are modelled in MCNP for a material containing Am-241 (see Figure 81 for an example energy spectra) and a material containing Cs-137 (see Figure 36 for example energy spectra). The new event sequence attempts to model a radioactive material traversing from Node01 to Node12 with anomalies occurring at 13:16:00 and 13:16:20, respectively. The source activity for the new radioactive materials were manually specified to provide a moderate probability of detection (i.e. moderate signal-to-noise) but remain small enough to remain near the normal background spectra to reframe from large spectral deviations. This is observed in Figure 94 showing the time-series count rate with the original anomalies from Event 3 highlighted in the light red regions and the inserted anomalous events highlighted in the light green regions on Node01 and Node12.

If the Bayesian framework proposed in this work was utilized for such scenarios, the time-series posterior probabilities can be calculated, as shown in Figure 95 and Figure 96 for the Am-241 and Cs-137 inserted sources, respectively.

Despite the relatively small source strength, the inserted anomalies at 13:16:00 and 13:16:20 on Node01 and Node12, respectively, are correctly classified as *different anomalies*, which again is primarily due to the large spectral deviation from both the normal background and initial alarm spectrum. To handle multiple, simultaneous anomalous events in a more formal manner, the Bayesian framework appends an additional hypothesis for the new anomalies and performs the identical posterior calculations as the three-hypothesis scheme. For example, if

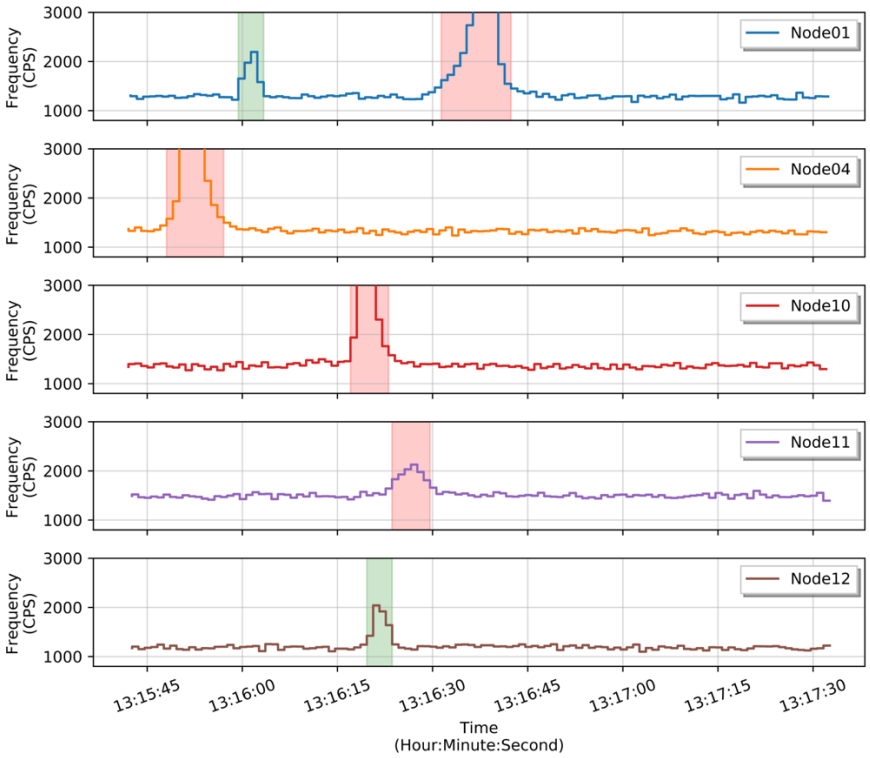


Figure 94: Event 2 time-series count rate with additional radioactive material present.

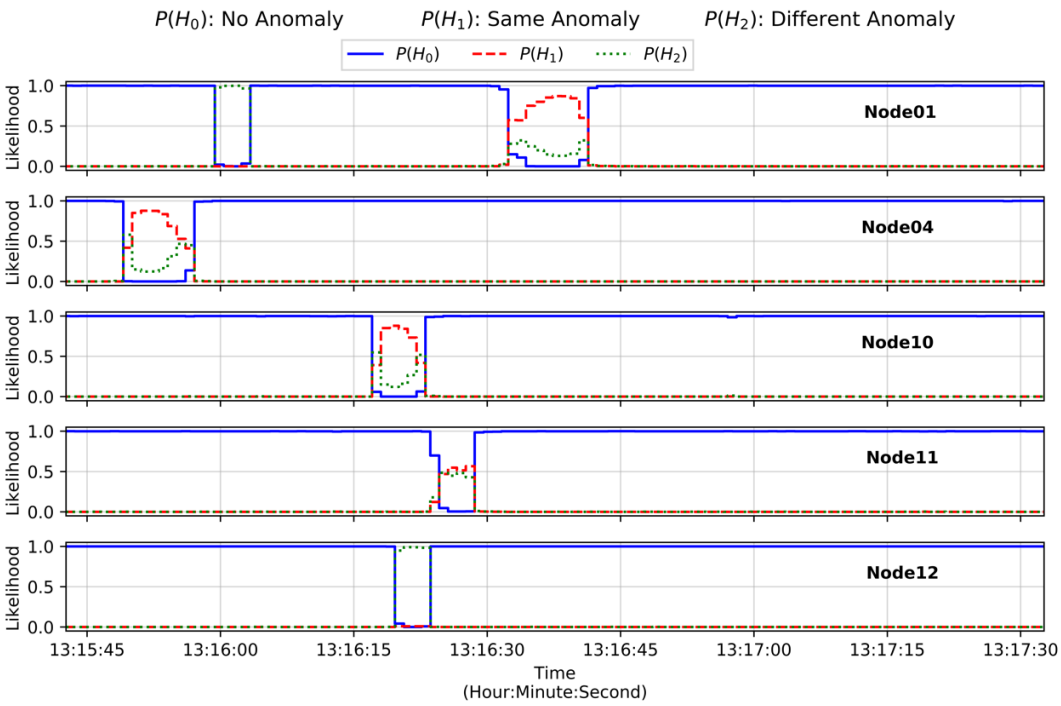


Figure 95: Event 2 time-series posterior probabilities with additional Am-241 source.

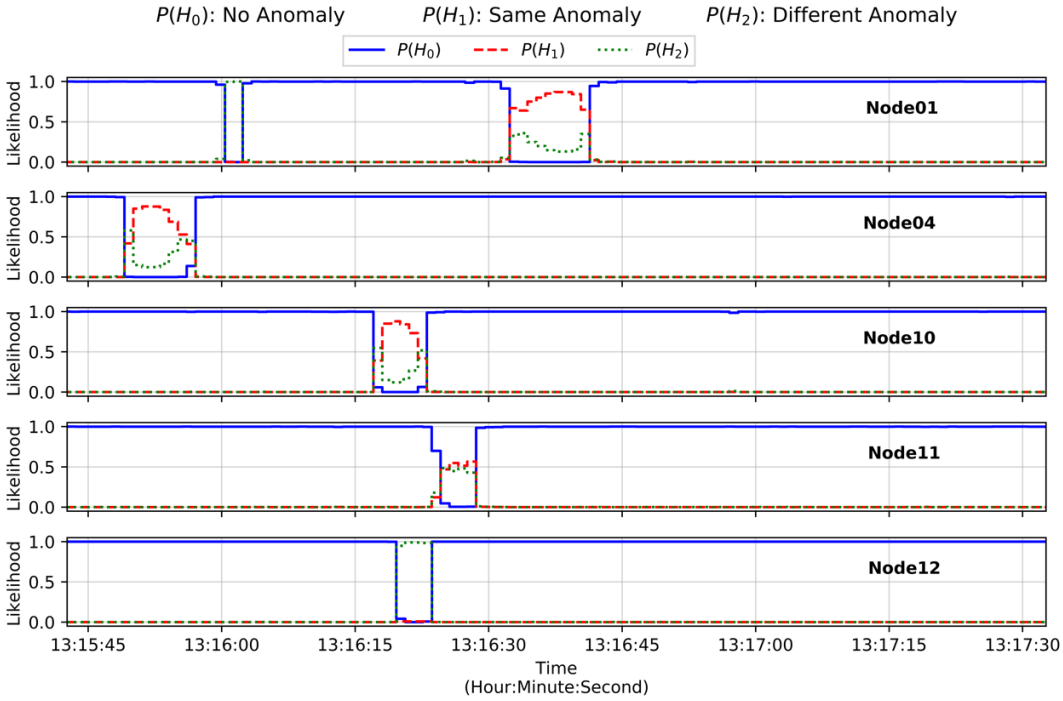


Figure 96: Event 2 time-series posterior probabilities with additional Cs-137 source.

the initial anomalous is referred to as *Event 0* and the Bayesian framework subsequently detects an anomaly at another node with a *different anomaly* classification, referred to here as *Event 1*, an additional hypothesis is appended to the hypothesis vector $\boldsymbol{H}$. The posterior probabilities would therefore be calculated by $P(H_j|\boldsymbol{X}^*) = \frac{P(\boldsymbol{X}^*|H_j)\,P(H_j)}{P(\boldsymbol{X}^*)}$. The value j represents the j^{th} index in the hypothesis vector $\boldsymbol{H}$ containing the likelihoods for the H_0: *no anomaly*, H_1: *Event 1 anomaly*, H_2: *different anomaly*, and H_3: *Event 2 anomaly*. Upon classifying a different anomaly, a new set of conditional probabilities are constructed using the extracted features, denoted by the asterisk superscript signifying the additional parameter to account for the H_3 hypothesis, $P(H_j|\boldsymbol{X}^*) = \frac{P(x_0^*|H_j)\,P(x_1^*|H_j)\,P(x_2^*|H_j)\;P(H_j)}{\sum_i P(x_0^*|H_i)\,P(x_1^*|H_i)\,P(x_2^*|H_i)\,P(H_i)}$.

In these example scenarios, the cosine similarity values range from 0.45 to 0.55 for both the Am-241 and Cs-137 sources relative to the initial anomalous spectra. The conditional probability $P(x_0^*|H_3)$ for the newly observed anomaly is obtained by creating a normal distribution with a mean obtained from the spectra classified as *different anomaly*. Figure 97 shows the unnormalized conditional probabilities from this method for the inserted source scenarios.

Identical methods to the previous events for calculating the expected time-of-arrival are employed for the multi-source scenarios and the conditional count rate probability $P(x_1|H_3)$ is identical to $P(x_1|H_1)$ and $P(x_1|H_2)$. The resulting time-series posterior for both the Am-241 and Cs-137 scenarios is provided in Figure 98 and Figure 99, respectively. Note that in the time-series analysis, the initial timestamp (i.e. 13:16:05) begins after Node01 classifies the inserted anomaly at 13:16:00 as $P(H_2)$ *different anomaly* and subsequently initiating four hypothesis methodology.

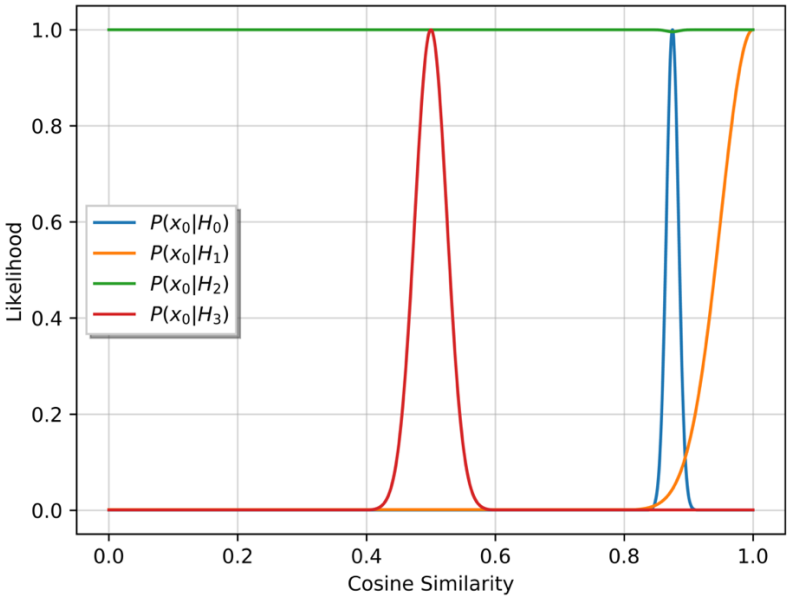


Figure 97: Event 2 cosine similarity conditional probability values for multiple source scenario.

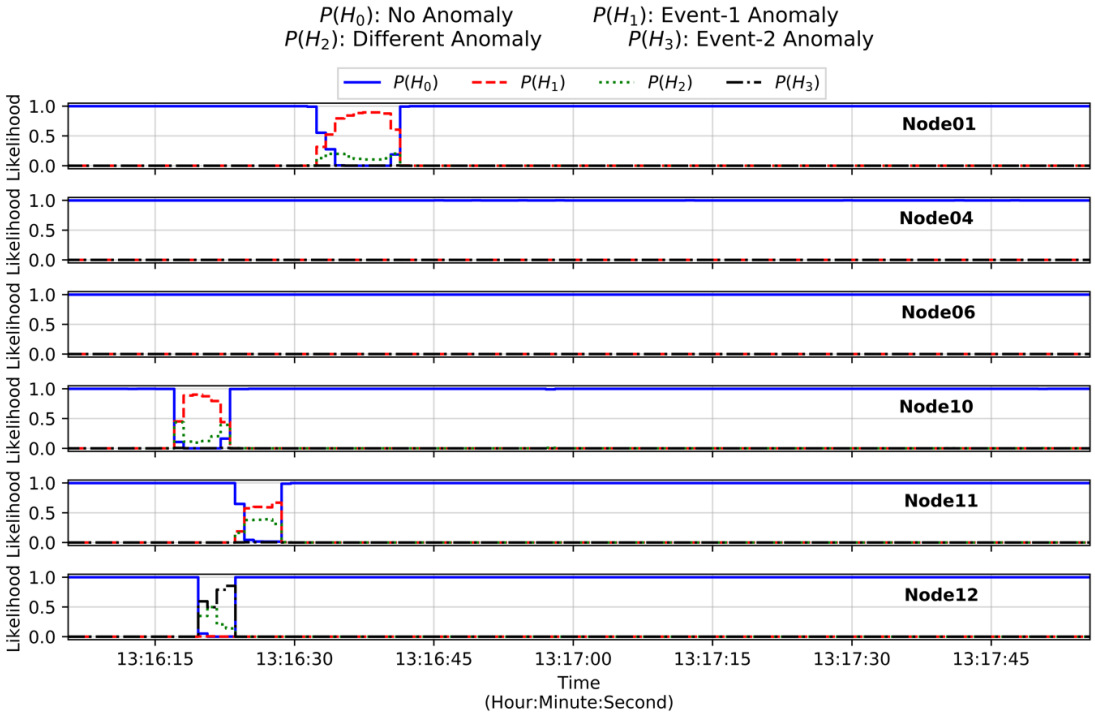


Figure 98: Event 2 time-series posterior probabilities with additional Am-241 source for multiple events.

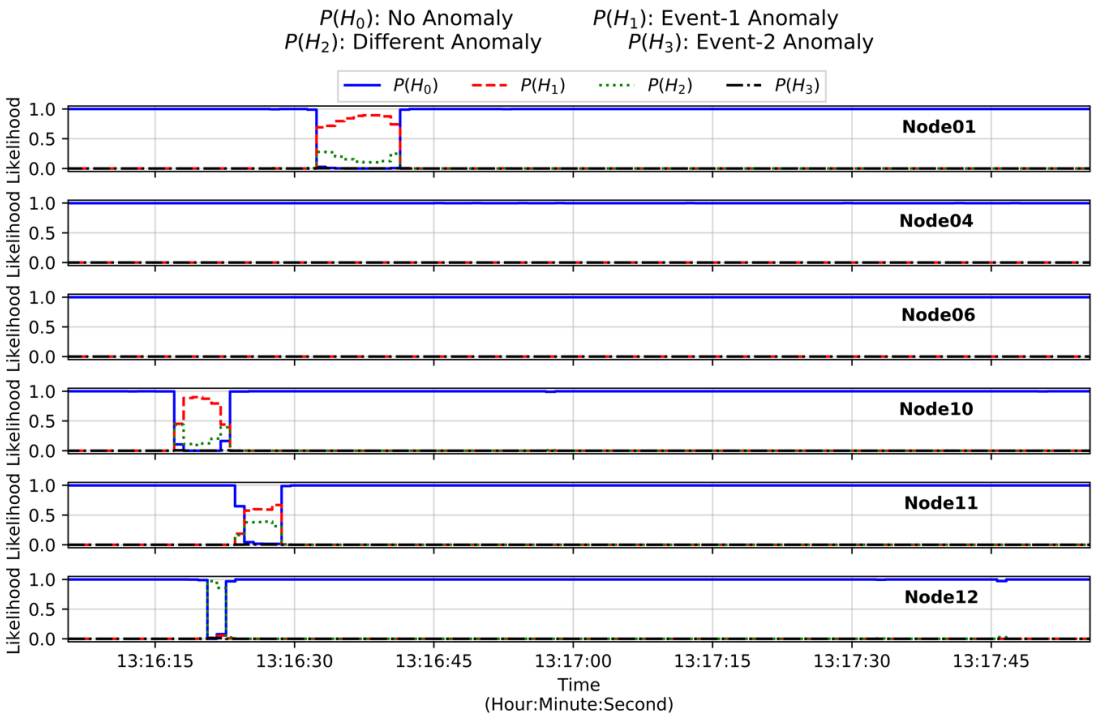


Figure 99: Event 2 time-series posterior probabilities with additional Cs-137 source for multiple events.

A few observations can be made from Figure 98 and Figure 99. First, recall that initial anomalous Event 2 sequence involves Node10, Node11, and Node01, in that order. The anomalies detected at said nodes correctly classified with the H_1: *Event 1 anomaly* remaining dominant. The anomaly from the inserted sources at 13:16:20 on Node12 is classified correctly as H_3: *Event2 anomaly* for the Am-241 source and incorrectly as H_2: *different anomaly* for the Cs-137 source. The cosine similarity values observed for the Cs-137 remained around 0.90 during the anomalous event, where the conditional probability $P(x_0|H_2)$ is dominant. This compounded by the short interaction time and a low count rate source, resulting in an energy spectrum more similar to the normal background spectra.

These results highlight the difficulty in correctly discriminating between each hypothesis utilizing these selected features in the Bayesian framework. Furthermore, it was stated previously that the Bayesian framework would not be able to correctly classify a different radioactive material with identical, or near-identical, photon emissions. In an attempt to improve the classification accuracy and simultaneously decrease the false-positive rate, other sensor modalities can be utilized to populate additional features in the data vector $\boldsymbol{X}$.

5: Distributed Multimodal Sensor Network

The previous the decision-making architectures can be expanded to incorporate previously derived outputs from the other contextual sensor modalities; detection attributes from the LIDAR point clouds and the video stream. The contextual sensors are first evaluated in their ability to supply additional information by inferring other parameters.

5.1: LIDAR Data Fusion

The output of the previously described LIDAR algorithm (see Section 3.2) may be utilized to infer additional parameters by fusing the radiation and LIDAR data. A similar concept was employed in previous works published at UF (see page 38) that correlated radiation count rate with multiple detected objects using cosine similarity to identify which detected object is carrying the anomalous material [76, 77, 78]. Recall that the cosine similarity compares the direction, or angle, between vectors, but the authors in [77, 78] utilize the method to compare the tracks of the detected object in a point cloud with the radiation count rate via the inverse-square law. In this work, multiple objects present simultaneously in the 2D LIDAR point cloud scene is improbable, thus attributing an anomaly to a specific object is not as important, but the process can be extended to approximate the activity of the radioactive material. Recall that the observed count rate (I) at the detector is impacted by a multitude of factors, but can be primarily thought to be affected by the solid angle (geometric efficiency, ϵ_{geo}), intrinsic efficiency (σ_{int}), and the distance between the material and detector (r), $I(S,r) = \frac{S}{4\pi r^2}\sigma_{geo}\sigma_{int}$. Of note, this method is overly simplified as no scatter component, shielding between the material and sensor, or energy-dependent intrinsic efficiency is incorporated. Since these attributes are neglected, the source intensity (S) can be thought of as more an *apparent* intensity and not a true source intensity. It is assumed that the 2” × 4” × 16” NaI(Tl) detectors have a relatively high intrinsic efficiency

across the emitted photon energies of interest, the intrinsic efficiency is assumed to be 1.0. Approximating the geometric efficiency is trickier, as the NaI(Tl) radiation detectors are right parallel piped resulting in a nonuniform solid angle as the observer rotates around the detector. The solid angle approximation provided previously may be used, but a more robust approach is to utilize particle transport models to simulate the detector response from radioactive sources at multiple locations around the radiation detector. This method was employed via MCNP using an accurate geometric representation of the node setup shown in Figure 100. A point source was positioned at distances 2, 3, 4, 5, and 6-meter distances from the center of the NaI(Tl) and rotated around the detected in 5° increments using the F2 surface current tally in MCNP. The output from the simulations is provided in Figure 101. For each degree within the 5° increments, the solid angle falls off at a rate of r^{-2} as expected and the solid angle is decreases and is flattened as the distance increases from 2 to 6 meters. Using this data along with the center points of the detected objects from the LIDAR point cloud algorithm to find the distance (r) node-to-detector distance, the *apparent* intensity (S) can be approximated from the observed count rate by linearly interpolating and extrapolating from the solid angle data.

The objective is to fit the inverse-square law relationship to the observed data with an additional background term added to the source dependent term, $I(S, r)$. An ordinary least-squares (OLS) fit, similar to method discussed in Section 4.2.2, is selected in this case, as it can be performed in quickly in real-time and adequate results were obtained during experimentation. An example MCNP simulation for point source moving in front of the radiation detector is provided in Figure 102, showing the background, source, and OLS fit to the data. Recall that the output from the OLS fit provides a mean and variance of the estimated parameter; that is, the *apparent* intensity.

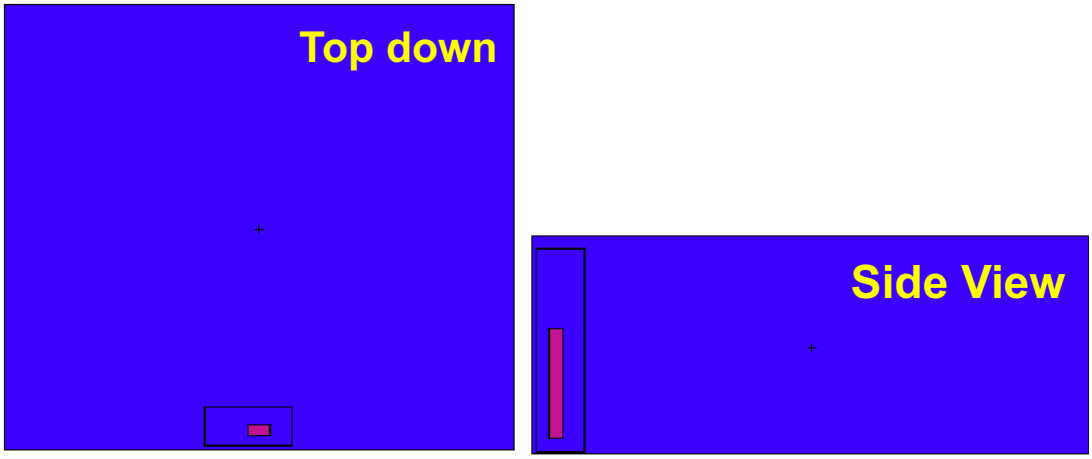


Figure 100: Geometry of MCNP model utilized in solid angle calculation.

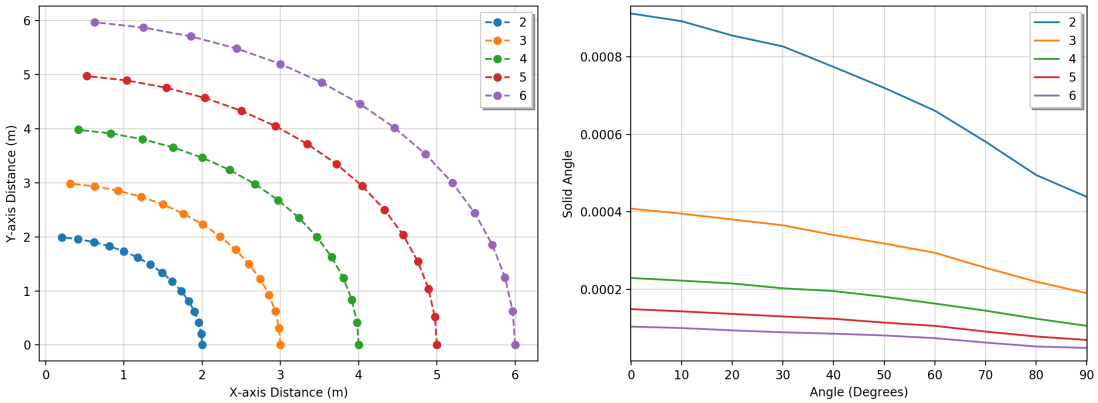


Figure 101: Location of point sources (*left*) and associated solid angle (*right*) from MCNP simulations from 2- to 6-meter radius from the sensor.

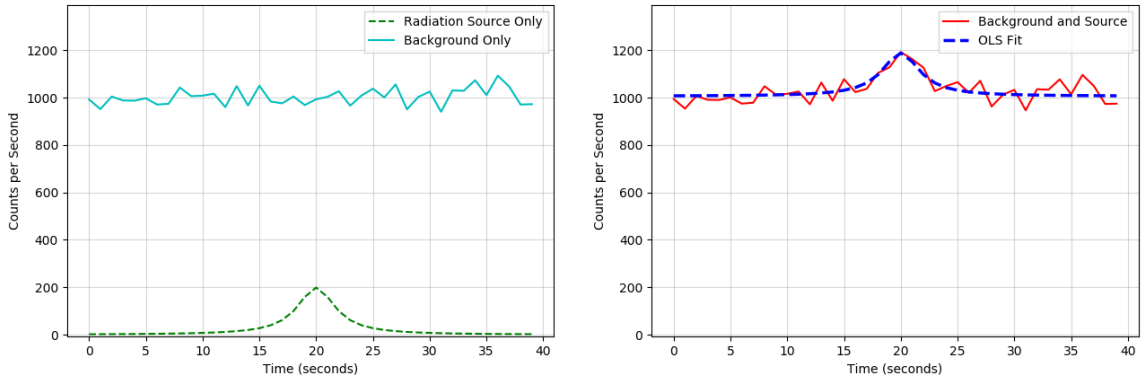


Figure 102: Ordinary least-squares fit to simulated count rate data for source intensity approximation.

This is a more efficient approach to creating new conditional probabilities and only requires for renormalization of the conditional distributions $P(x_1|\boldsymbol{H})$ after updating the *same anomaly* probability, $P(x_1|H_1)$. To provide additional context, assume the data from Figure 102 is the count rate data observed during the initial anomaly and results in an *apparent* intensity of 200 ± 75 counts per second using OLS. Using that information to create an updated conditional probability $P(x_1|H_1)$, the resulting conditional probabilities $P(x_1|\boldsymbol{H})$ are shown in Figure 103. The primary difference is shown in the H_1 *same anomaly* that constrains the $P(x_1|H_1)$ to a small area via the created Gaussian distribution. Unfortunately, this does not provide an improvement in the ability to differentiate between H_1 *same anomaly* and H_2 *new anomaly* as the conditional probability $P(x_1|H_2)$ is dominate in the majority of the count rate region above background.

When evaluating this methodology for tracking anomalies, the accuracy of the decision-making architecture suffered when the apparent intensity observed for the same radiological material at subsequent nodes was drastically different than the initial apparent intensity. This did not occur when strong signal-to-noise ratios were measured in the radiation count rate data, but occurred quite frequently for low signal-to-noise ratio scenarios, such as detected objects either at great distances from the detector or lack of multiple samples from an detect objects moving quickly. To create a more robust methodology, the approximate apparent activity from the initial is instead used to approximate the expected count rate at subsequent nodes based on all detected objects in the LIDAR data. This should increase the accuracy of the classification process in low signal-to-noise ratio scenarios by calculating the expected count rate contribution governed primarily by the inverse-square law. When utilizing this method, the conditional probability $P(x_1|H_1)$ for *same anomaly* is required to be updated dynamically as new object position data becomes available. The probability distribution is calculated in an identical manner to the

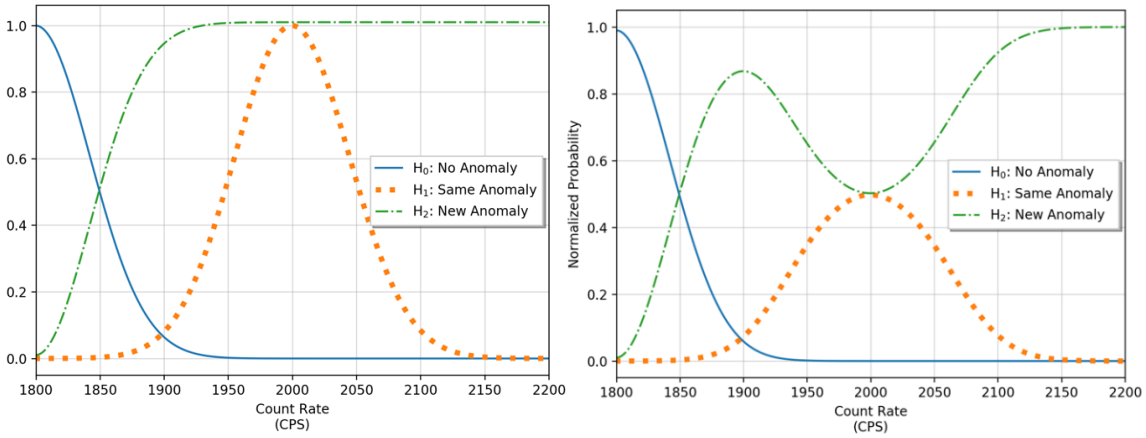


Figure 103: Updated unnormalized (left) and normalized (right) count rate conditional probabilities for arbitrary count rate data.

previous approach using the inverse-square law with the background distribution added. For example, let's examine a scenario depicted in Figure 104 where a vehicle carrying a radiological material (i.e. point source in this case) with an activity of approximately 4500 counts per second and a normally distributed background with a mean of 1000 counts per second.

Using the inverse-square law along with the geometric efficiency, the expected count rate is estimated and the corresponding conditional probability $P(x_1|H_1)$ is created at each time step, as depicted in Figure 105. Using an estimated apparent intensity of 4500 ± 100 counts per second from the anomaly (i.e. randomly selected in this case), $P(x_1|H_1)$ likelihood initially is similar to that of the $P(x_1|H_0)$ *no anomaly* likelihood as the expected contribution to the count rate remains low. As the vehicle approach the detector, the solid angle will increase and the expected observed count rate for H_1 *same anomaly* also increases in the count rate domain. In this example, the vehicle traverses directly by the detector and provides a relatively strong signal-to-noise ratio as observed in time-series data in Figure 106. If the vehicle diverts its route prior to passing the detector, the variance in the OLS fit would increase and the observed count rate spike would be minimized. Thus, this technique should provide a more robust methodology for estimating the conditional count rate probabilities $P(x_1|\boldsymbol{H})$, but at the cost of computation power that is required to update the probability space at each time step.

The LIDAR data was further evaluated in an attempt to provide an object classification based on the extracted data from the clustering algorithm. It was hypothesized that analyzing certain features from the LIDAR data would provide an ability to differentiate between classes (e.g., human, car, truck). If an object classification were possible, the system could fuse the classification data between the LIDAR and video data streams to improve the confidence in the object classification, or, even better option, reduce the reliance of the video classification

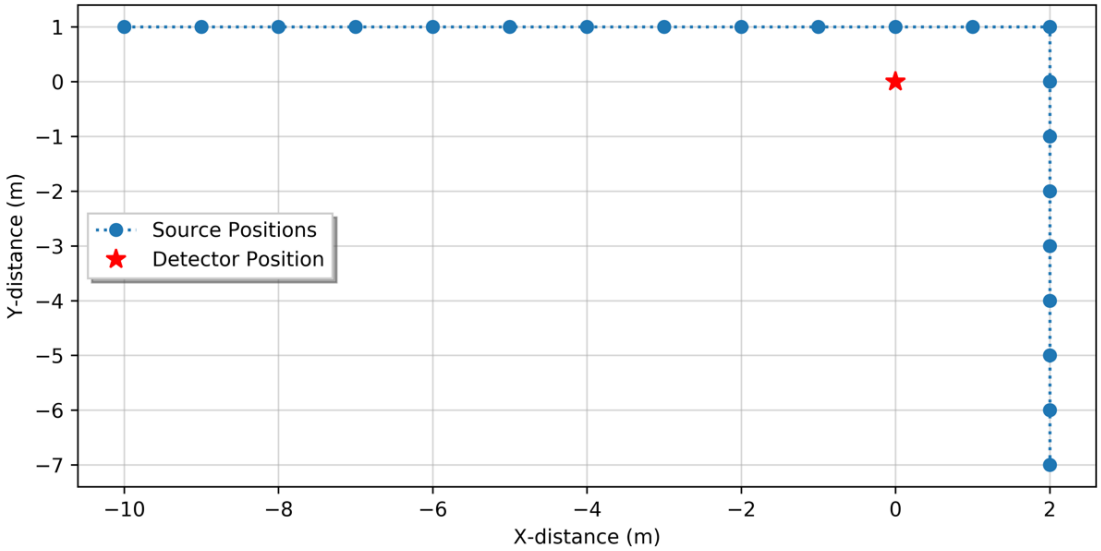


Figure 104: Example scenario utilized for apparent intensity approximation.

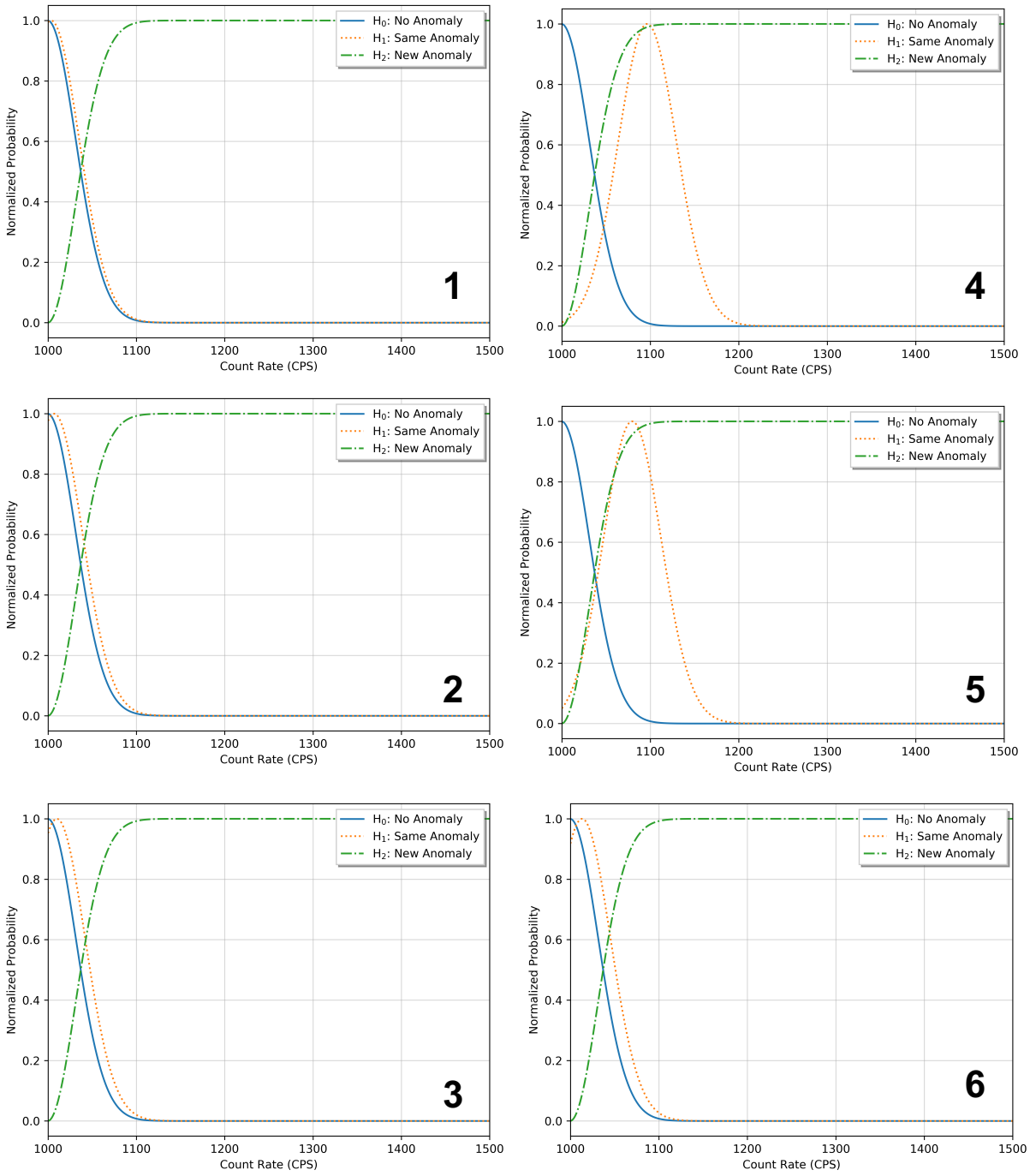


Figure 105: Dynamically updated count rate conditional probabilities for example scenario.

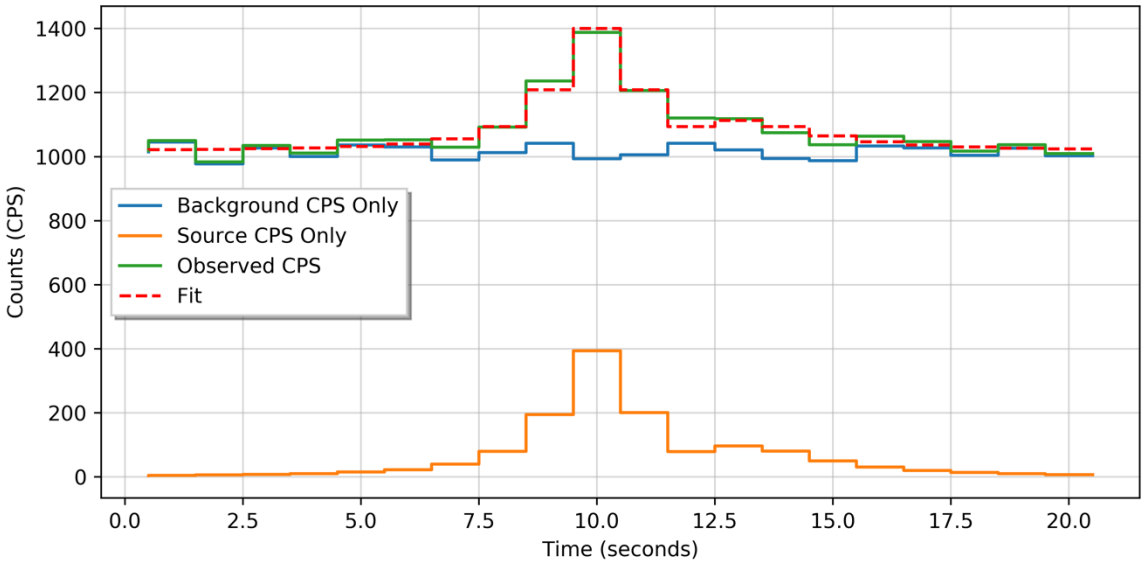


Figure 106: Count rate and OLS fit for example scenario.

algorithms by providing high confidence level classifications. The list of features extracted from the detected clusters in the 2D point clouds via the occupancy grid algorithms include the time of detection, position of cluster, cluster length, cluster width, and cluster velocity.

To limit the time to manually label the detected anomalies in the LIDAR data, the video data is utilized to create pseudo-classifications of the objects. It is assumed in this work that the object classifications from the video algorithms are of high accuracy, thus the initial classifications for detected objects in the LIDAR data with near-identical timestamps (i.e. within 3 seconds difference between the video and LIDAR timestamps) are assumed to be of high accuracy. An example of the pseudo-classifications, Figure 107 provides the median velocity from the Kalman filter along with the median cluster area from LIDAR data for each detected event observed at one-week at Node12. Across all the nodes, approximately 70% of the pseudo-classifications remain classified as *Unknown* as the timestamps from the detected objects in the LIDAR data are not within 3 seconds of any video classifications. Assuming the time is correctly synchronized across the data streams, the amount of *Unknown* classifications is contributed to an object being detected in the LIDAR data via the occupancy grid algorithm and not entering the field-of-view of the video data. Absent of proper labelled dataset, the use of the LIDAR data to perform object classification, the LIDAR data is not utilized in this work for object classification. The author notes that utilizing poorly labelled data or data without labels for training classification algorithms is possible via unsupervised algorithms, such as *k*-means clustering, DBSCAN, or autoencoders, which looks for underlying structures in the data itself [139].

5.2: Video Data Fusion

Data fusion between video data and other data streams has observed increased popularity in recent years due to the increased computational power available to process the feature-rich data

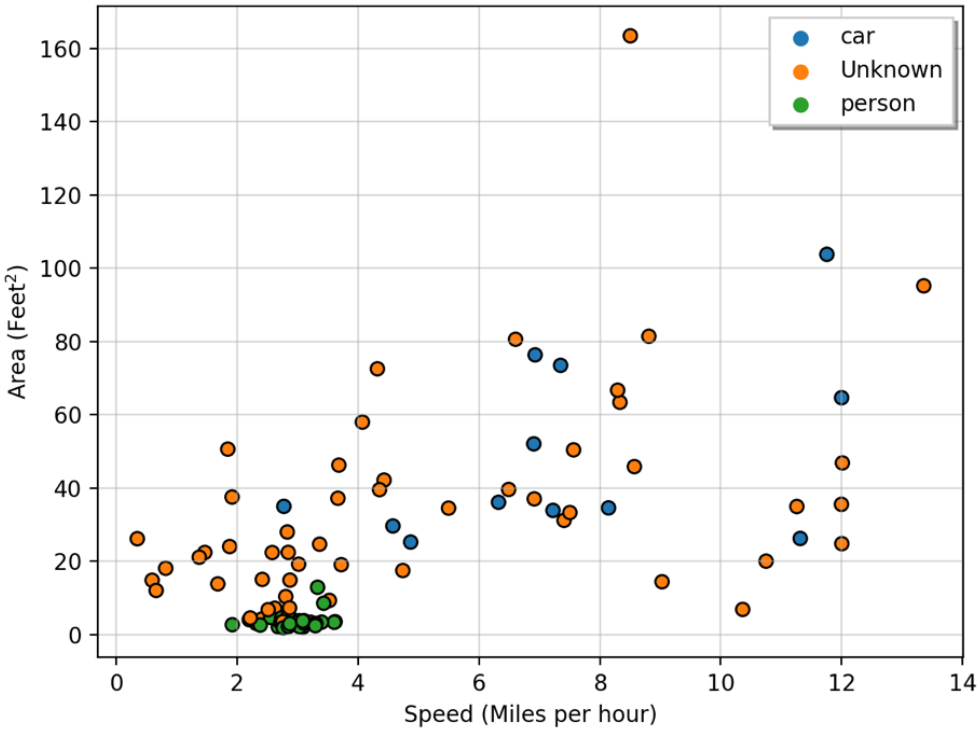


Figure 107: Example extracted features from LIDAR data with pseudo-classifications.

in real-time. In this work, the primary reason to fuse data is to increase the ability to detect and track a radiation anomaly. This involves utilizing the features extracted (i.e. object classification) from the video data with the radiation signal. It has been well-documented that objects, primarily large vehicles, in the environment surrounding the radiation detectors will suppress the terrestrial background resulting in a decreased count rate, often times referred to as a shadow effect. This scenario is more common when vehicles are moving slowly near a radiation detector, such as radiation portal monitors positioned at border crossings to monitor traffic [165, 166, 167, 168]. The degree of suppression, or shadow, is primarily dependent on the density of the vehicle, the velocity of the vehicle, the distance of closest approach, and the background source spatial distribution surrounding the detector. This additional information could modify the expected observed count rate for H_0: *no anomaly* given an object is observed in scene which may increase the ability the discriminate between hypotheses in the count rate data.

It was hypothesized that the degree of suppression would be unique for each type of object detected by the video processing algorithm (e.g., person, car, or truck). Prior to the ORNL sensor array being deployed, a controlled experiment was organized to better understand the effect that vehicles have on the collected radiation signal in a more controlled environment [169, 170]. A 2" x 4" x 16" NaI(Tl) radiation detector, identical to the deployed detectors in the nodes in the ORNL sensor array, was positioned at the edge of the roadway on the ORNL reservation to measure the background radiation for different objects moving through the scene. A truck and a human surrogate made of two plastic buckets stacked full of water (i.e. 6' in height and 12" diameter) were positioned in the road at 3 meter and 1 meter away from the sensor for the truck and human surrogate, respectively. The objects were stepped through the scene in front of the sensor, as depicted in Figure 108, and

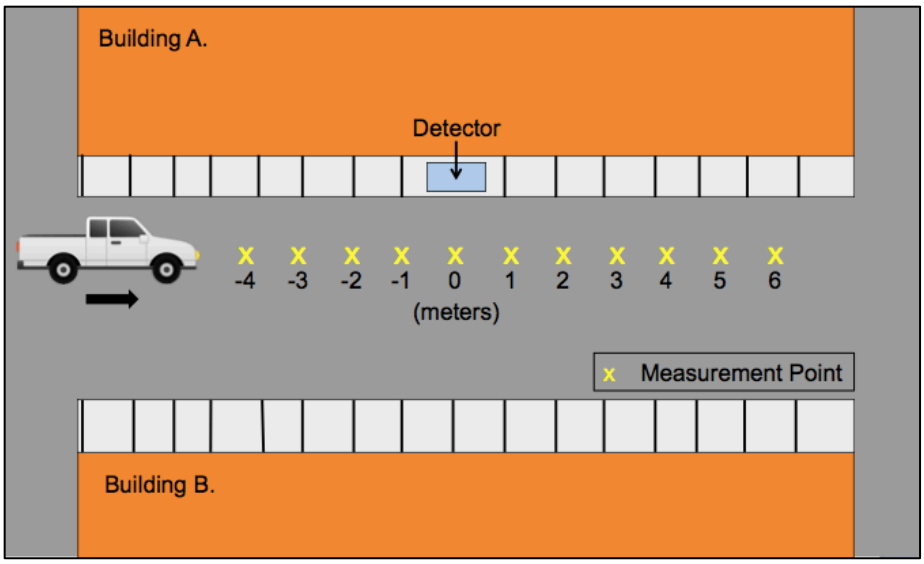


Figure 108: Illustration of data collection between buildings for controlled suppression study.

a 25-minute static measurement was collected at each location specified in the Figure 108. The mean background count rate and the measured mean count rate at each position are provided in Figure 109 with 2σ error bars. The truck and human surrogate deviate from the measured mean count rate by different amounts. The rather unique suppression the truck occurs after the zero relative-position, as visualized in Figure 109, is due to the measurement using the front of the truck as the positioning and not the center. This results the largest deviation, or residual, from the mean background count rate after the zero relative-position as the center of the truck passes with the detection position, increasing the solid angle subtended by the truck. The human surrogate observes a very slight deviation from the man count rate, which is attributed to the low-Z value and density of the materials (i.e. plastic and water), decreasing the attenuation of the background radiation.

The observed deviations for the long-dwell static suppression experiments could be used to update the decision-making architecture by changing the existing architecture to incorporate the conditional probability, such as $P(x_1, x_4|\boldsymbol{H})$ where x_4 represents the object classification from the video feed, but a more realistic experiment should be performed to identify the impact on objects moving through the environment. Recall that while an object moving through the scene, a finite number of samples are collected based on the timing resolution of the data acquisition system. Furthermore, the objects move during the sampled time and thus it can be difficult to accuracy quantify the deviations in the observed data. An additional experiment with an identical setup was performed for a semi-controlled environment where roadway was open to traffic travelling in one direction around 2 to 3 meters from the detector system [170]. Based on the data observed from the static experiment, the expected deviations will be largest within 5 meters on either side of the detector. Using this information in the semi-controlled experiment,

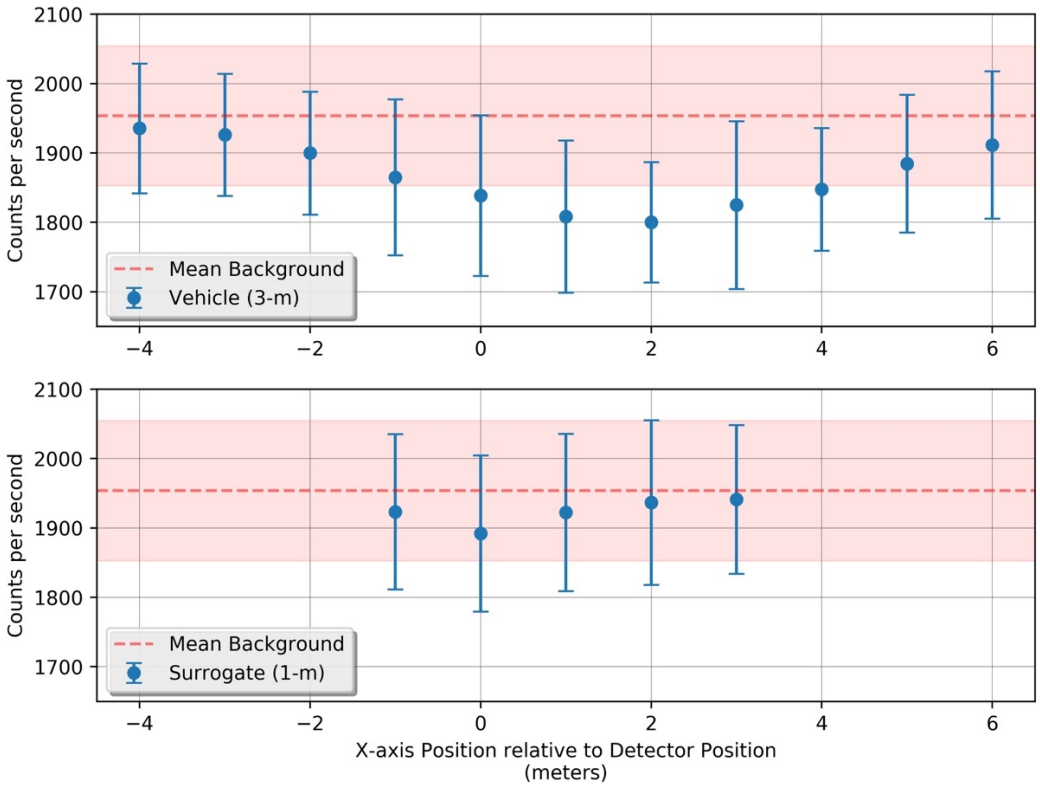


Figure 109: Measured count rate for truck and human surrogate long-dwell controlled suppression study.

the acquisition system collects count data only when the LIDAR unit detects an object (designated as the term *clutter* in this instance) in the roadway within 5 meters from either side of the detector, a depicted in Figure 110. In Figure 111, the background and motion event count rate distribution is provided showing a minute suppression from background for the motion events. Approximately a 7% difference in the mean count rates from the background to motion event was observed (i.e. 180.1 ± 13.8 and 193.9 ± 15.8 counts per 100-milliseconds for the background, respectively). The amount of suppression is primarily due to a large majority of the background count rate originating from a large, multi-story brick building across the street from the detector. Thus, as the object move between the building and detector, a larger portion of the background signal can be absorbed or scattered away from the detector. The amount of suppression is primarily due to a large majority of the background count rate originating from a large, multi-story brick building across the street from the detector (i.e. recall the background source distribution is a key parameter). Thus, as the object move between the building and detector, a larger portion of the background signal can be absorbed or scattered away from the detector.

Once the ORNL sensor array came online, the data collected from Node12 was utilized to evaluate if any trends or strong correlations existed similar to the controlled and semi-controlled studies. Node12 was selected as the preferred location as the system is located directly beside the road (i.e. ~1 foot from the edge of the roadway). Using the LIDAR and video processing algorithms discussed in previous sections of this report, one week of data was analyzed to quantify the impact each object classification induces on the count rate. The collected data are provided in Figure 112 for each category (e.g., truck, person, car, and background) with no large deviations from background. The person category is particularly of unattractive due to only a few

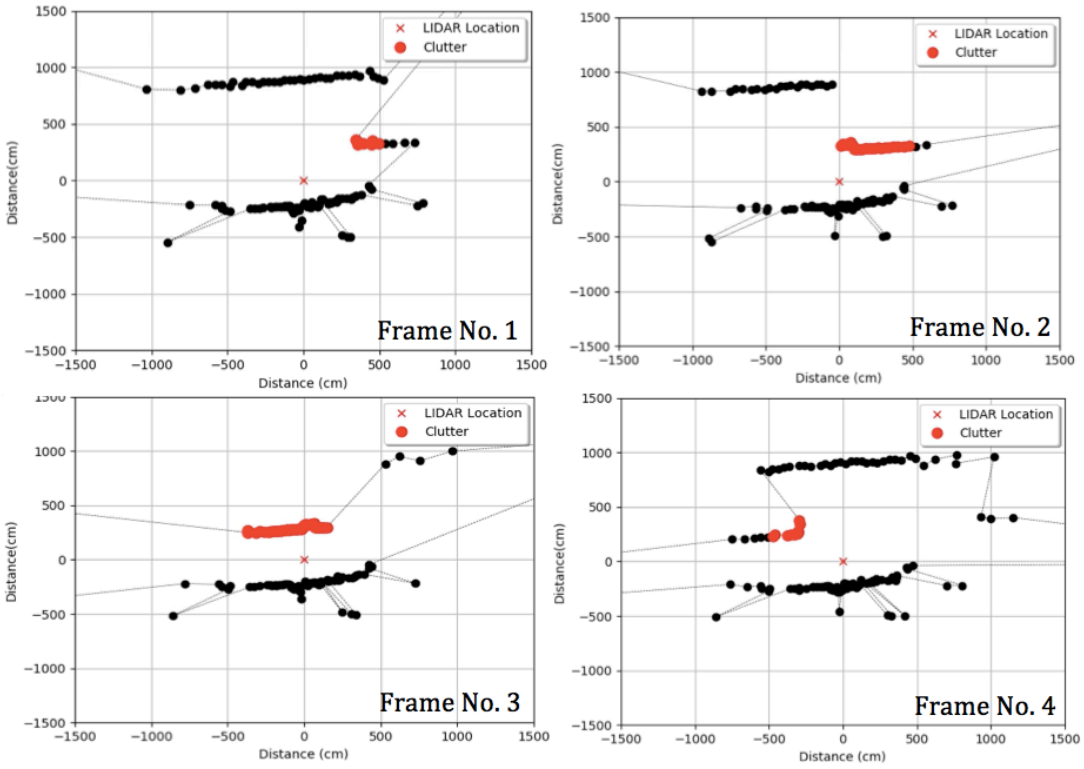


Figure 110: Example LIDAR output for semi-controlled environment suppression study.

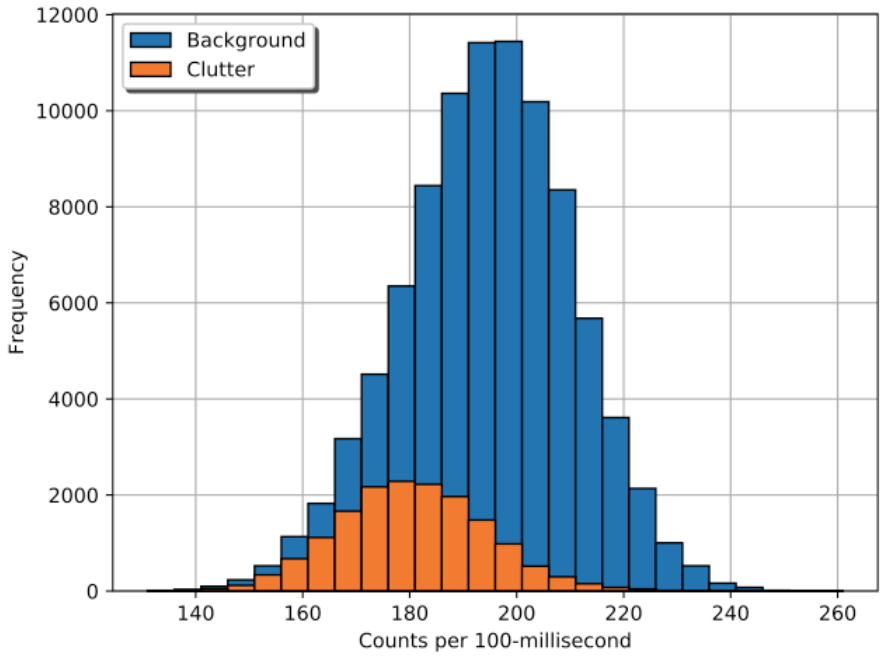


Figure 111: Distribution of collected count rates for semi-controlled environment suppression study.

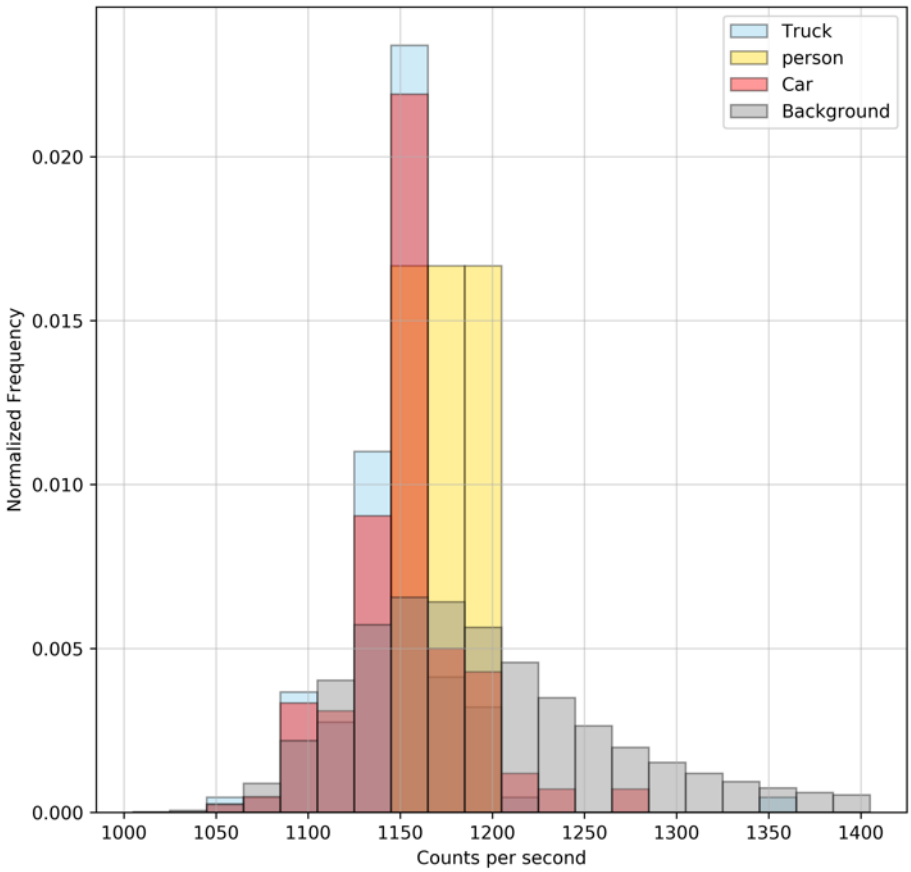


Figure 112: Distribution of collected count rates for uncontrolled environment suppression study.

events observed providing only a small sample of the actual distribution, even if the expected deviation is low. The truck and car categories have near-identical distributions, but both have no strong deviation from the background distribution. This was expected for a few reasons. First, Node12 is positioned away from any large structure and the majority of the background radiation is thought to originate from the ground surrounding the detector. Therefore, the count rate suppression should be less, absent of any dominant background source on the opposite side.

Second, the methodology used to initiate the collection of a motion event was near identical to the semi-controlled environment, but the objects were not constrained by speed or distance of closest approach, contrary to the previous two experiments. These attributes force additional noise into our measurement as the data acquisition system may be only be capable of a few samples of a motion event and the object may be further away from the detector. These contribute to the noise in the distributions. For these reasons, the distribution deviates only slightly from the measured background distribution and fusing this information into the decision-making architecture is not utilized in this work as it is assumed to provides only modest, if any, benefit to the architecture's detection accuracy.

Absent of any approach to fuse the radiation data with the video classification data, the video data can still be directly incorporated in the decision-making architecture for tracking anomalies. The objective is to compare the initial object classified within the subsequent objects detected in other data streams. This subject has been observed increased popularity in recent years, much like most analytic tools for video data, and is often referred to as object re-identification. Most notably, the NVIDIA Corporation has sponsored the AI City Challenge over the last several years to address city-scale multi-camera vehicle tracking and re-identification using both synthetic and real-word training data [172, 173, 174]. These techniques utilize various

architectures that used neural networks to compare different views of objects along with extracted features to re-identify vehicles in an area. The current approaches employed to solve this problem remain in the early TRL stages but could provide a massive benefit for vehicle tracking in an urban environment in the very near future.

This work takes a relatively straightforward approach to re-identification, as the other orthogonal sensor modalities (e.g., LIDAR or radiation sensors) can aid in the re-identification process for tracking an object attributed to a radiation anomaly. The object classification data obtained from the final layer of the YOLOv3 network can be thought of as a discrete distribution, as the classes in the final layer are discrete variables even though the probability values predicted for each classification can be continuous. Rather than only transmitting only the selected classification from the three labels to other nodes in the network, the final layer of the YOLOv3 can be transferred and used to build a discrete probability distribution. For example, if the average values from the final output layers during the event are [0.21, 0.95, 0.62] for [person, truck, car] as depicted in Figure 113, the discrete values can be used to represent the belief or likelihood of the object belonging to one of the categories. In the Bayesian context, this can be directly inserted by representing the discrete values as probability, $P(x_3|\boldsymbol{H})$, where the object detected in the initial anomaly ($x_{3,i}$) will represent H_1 *same anomaly* via the conditional probability $P(x_{3,i}|H_1)$. Unfortunately, the H_0 *no anomaly* and H_2 *different anomaly* can exist for any and all combinations of detected objects and thus should be equal to 1.0 for all objects.

This process is only employed with an object is detected in the video data, as any other time would result in an empty conditional probability $P(x_3|\boldsymbol{H})$, as a result during times with no detected objects, the original $P(\boldsymbol{H}|x_0, x_1, x_2)$ construct is used. Furthermore, this construct is not

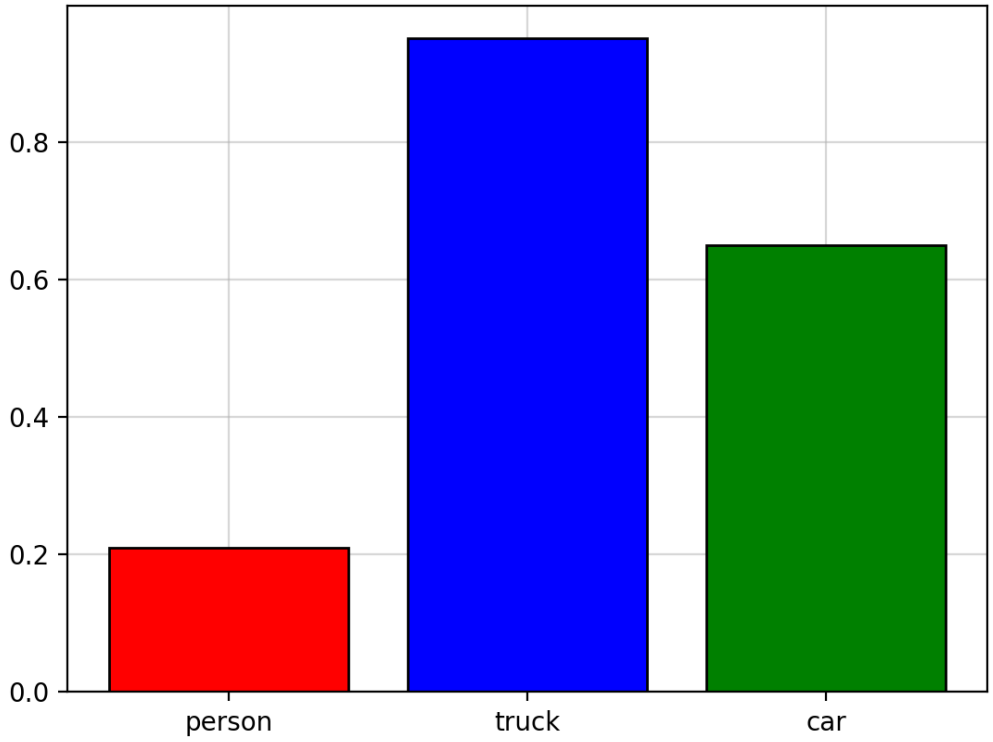


Figure 113: Example discrete distribution for arbitrary video classification output.

ideal as it provide no additional value to discriminating between the hypotheses. Nevertheless, this construct will be utilized to understand the implication of fusing the data this manner.

5.3: Multimodal Bayesian Classification Methodology

Using the previous analysis to dynamically update the $P(x_1|H_1)$ probability by fusing the LIDAR data and the addition of a new discrete probability distribution via the video classification data, the Bayesian decision-making architecture can easily be updated to incorporate the new information. The modified $P(x_1|H_1)$ can be incorporated in a straightforward manner by replacing the $P(x_1|H_1)$ and dynamically updating the conditional probability and the new conditional probability $P(x_3|\boldsymbol{H})$ is easily incorporated by appending the probability to the current posterior calculation as shown $P\left(H_j\middle|\, x_0, x_1, x_2, x_3\right) = \frac{P\left(x_0|H_j\right) P\left(x_1|H_j\right) P\left(x_2|H_j\right) P\left(x_3|H_j\right) P(H_j)}{\sum_i P(x_0|H_i) P(x_1|H_i) P(x_2|H_i) P(x_3|H_i) P(H_i)}$. Using this updated Bayesian formula, the previous three example events from the ORNL sensor array are analyzed using the radiation, LIDAR, and Video data. Recall that in Event 1, the initial anomaly was observed at Node06 with subsequent anomalies measured at Node04, Node10, and Node01 Each sensor modality (i.e. radiation, LIDAR, and video) at each alarm node successfully detected an anomaly in their respective data stream. The resulting time-series posterior for the event is provided in Figure 114.

It can be observed when comparing the original posterior values in Figure 92 to update values in Figure 114 that the H_1 *same anomaly* classification increases and is dominant during the event. In attempt to compare the two, the average likelihood values calculated for each posterior during the anomalous event are calculated and provided in Table 10.

It should be noted that using the average of the posterior probabilities during the event may provide bias results as the instantaneous probabilities at each time-step may be more important to an end-user, rather than averaged over the event. For example, the anomalous

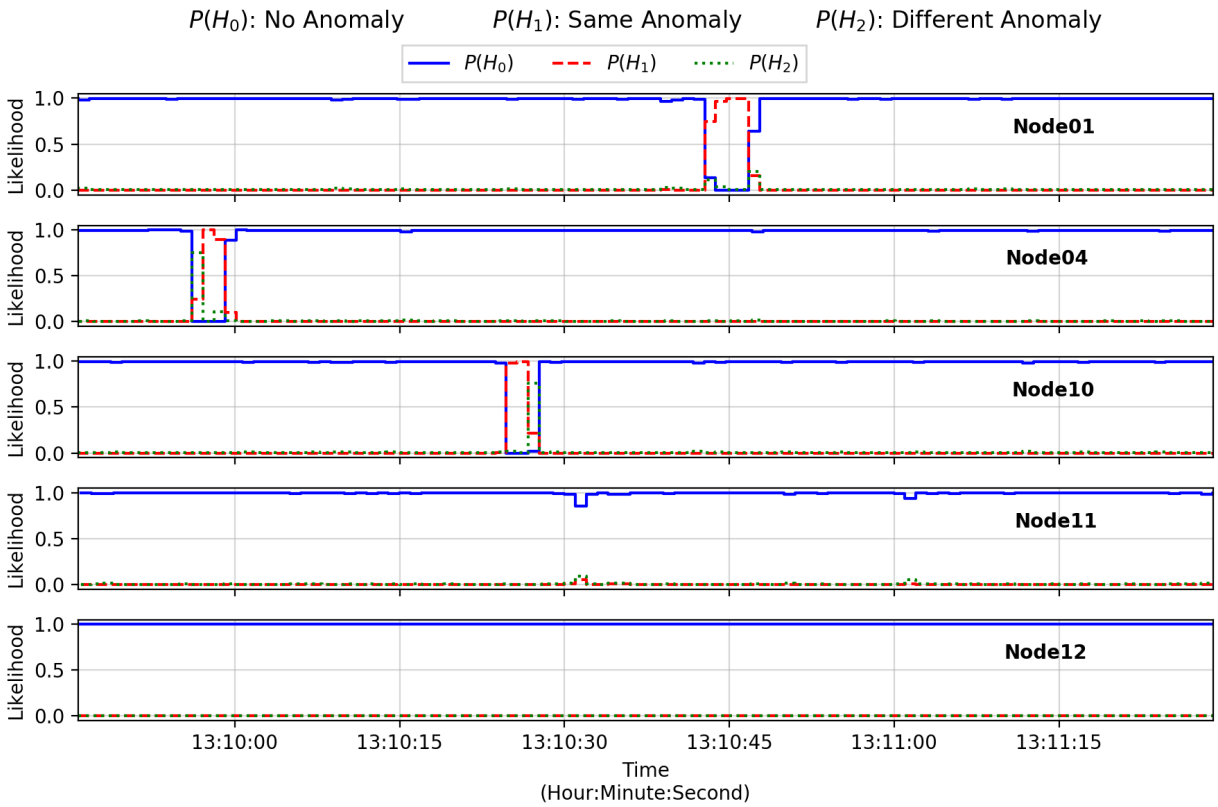


Figure 114: Event 1 time-series of the posterior probabilities using the multimodality Bayesian framework.

Table 10: Event 1 average Bayesian network output comparison during anomalous event.

		Radiation Network	**Multimodality Network**
Node01	H_0	7.5%	1.7%
	H_1	**57.3%**	**90.4%**
	H_2	35.2%	7.9%
Node04	H_0	15.2%	0.1%
	H_1	**54.3%**	**68.3%**
	H_2	30.5%	31.6%
Node10	H_0	15.2%	1.5%
	H_1	**54.3%**	**69.9%**
	H_2	30.5%	28.6%

events detected at Node04 and Node10 in Event 1 observe the H_2 *different anomaly* classification being the most dominant, or MAP. Therefore, if the instantaneous results are utilized to classify the system state, a misclassification would occur at the beginning of the event on Node04 and at the end of the event on Node10, as shown in Figure 114. Additionally, providing the average posteriors probabilities is dependent on the start and end times selected for the event. In this case, it is considered an event when the H_0 *no anomaly* classification is less than either H_1 or H_2, but is bias towards the decision-making architecture utilized in this work. If a more advanced algorithm, for instance, were to detect the radiation anomaly at an early time, a direct classification comparison is not advisable as the more advanced algorithm was able to detect the anomaly earlier. In lieu of any alternative methods to compare the radiation-only Bayesian framework and the multimodalilty Bayesian framework, the averaged posterior probabilities are used to equate the two.

Event 2 observed an initial anomaly at Node06 with subsequently anomalies observed at Node04, Node10, Node11, and Node01. The results from the original, radiation-only sensor network (see Figure 93) provided no misclassifications, but an increased competition between H_1 *same anomaly* and H_2 *different anomaly* which was attributed to the calculated cosine similarity values residing in the overlap region between the two conditional probabilities, $P(x_2|H_1)$ and $P(x_2|H_2)$. The multimodal results, shown in Figure 115, provide a higher posterior probability values for H_1 *same anomaly* and a suppression of H_2 *different anomaly* posterior values across all four anomalous event detected. This leads to an enhanced average posterior value for each anomalous event as provided in Table 11.

Similar to the alarm path in Event 1, Event 3 observed an initial anomaly at Node06 with subsequently anomalies observed at Node04, Node10, and Node01, respectively. Each modality

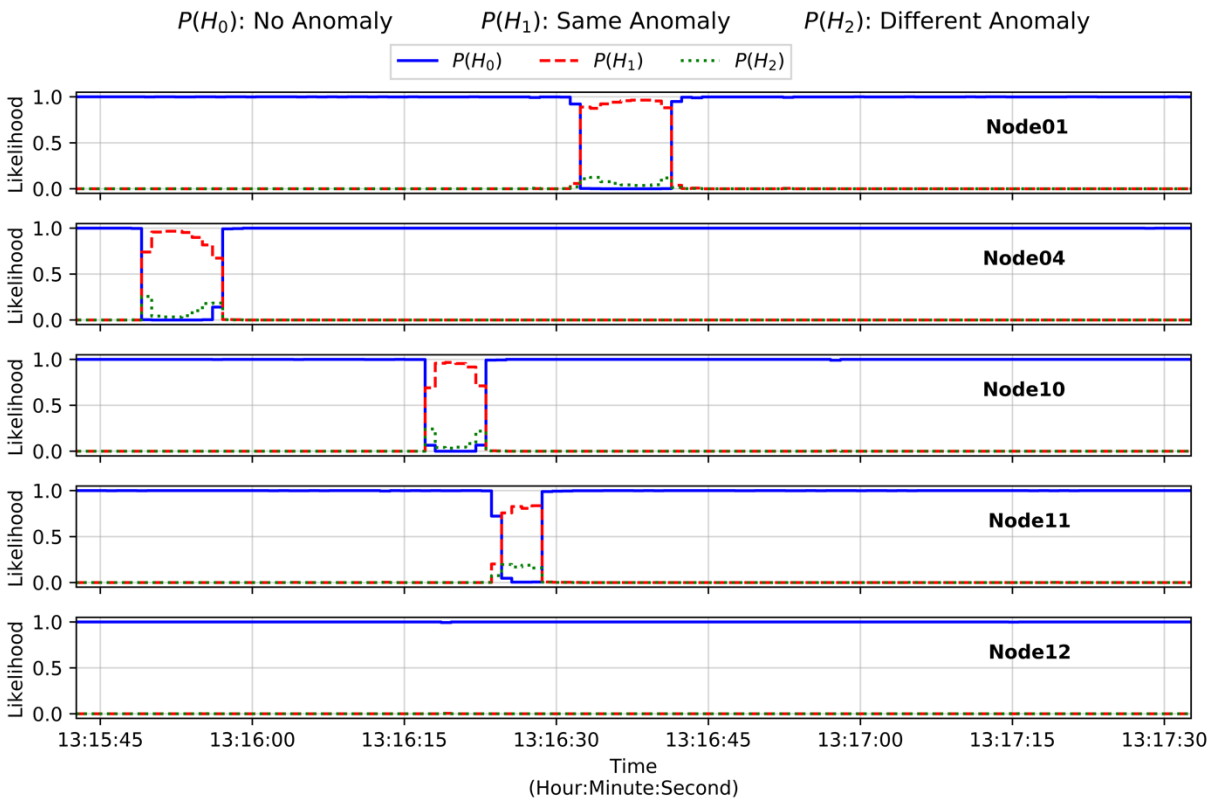


Figure 115: Event 2 time-series of the posterior probabilities using the multimodality Bayesian framework.

Table 11: Event 2 average Bayesian network output comparison during anomalous event.

		Radiation Network	**Multimodality Network**
Node01	H_0	0.2%	0.4%
	H_1	**77.1%**	**90.7%**
	H_2	22.7%	8.9%
Node04	H_0	0.2%	2.6%
	H_1	**77.5%**	**83.7%**
	H_2	22.3%	13.7%
Node10	H_0	0.1%	3.2%
	H_1	**82.6%**	**83.1%**
	H_2	17.3%	13.7%
Node11	H_0	1.5%	2.3%
	H_1	**53.7%**	**75.6%**
	H_2	44.8%	22.2%

at each alarm node successfully detected an anomaly in their respective data stream. The resulting time-series posterior for the event is provided in Figure 116. In comparison to the posterior probability values from the radiation-only sensor network (see Figure 91), the multimodality posterior probability value show an increased values in H_1 *same anomaly* with average values compared in Table 11. Similar to the results from Event 1, the average posterior values show an enhanced classification, but misclassifications can be made if a classifier method that utilizes time-step approach for making decision. Fortunately in this event, the misclassification for H_2 *different anomaly* remains suppressed, but as the count rate and consequently, the cosine similarity values decrease, the H_0 *no anomaly* becomes dominant, which is expected. Overall, however, the detection of the anomalies observes an increased confidence.

In all three events, the MAP classification did not change and increased the correct confidence in the correct classification, namely *same anomaly*, while suppressing other the other hypotheses, or classification categories.

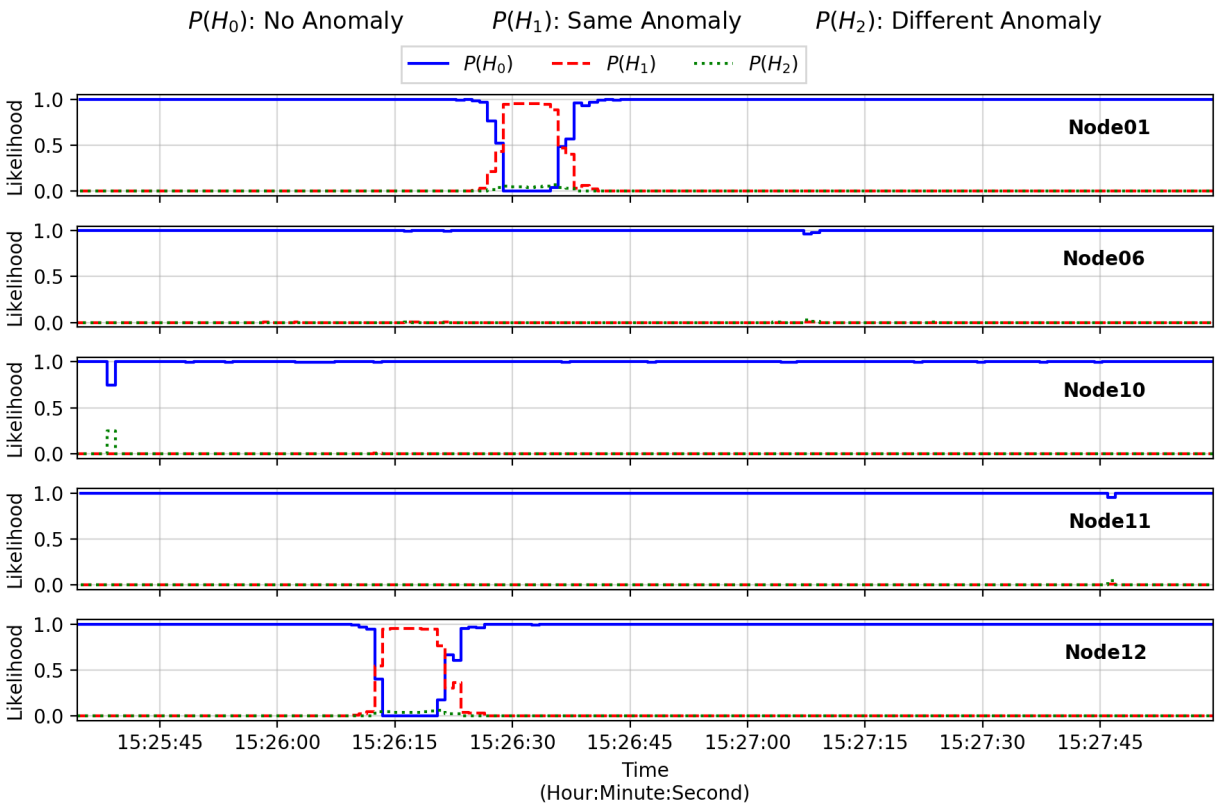


Figure 116: Event 3 time-series of the posterior probabilities using the multimodality Bayesian framework.

Table 12: Event 3 average Bayesian network output comparison during anomalous event.

		Radiation Network	**Multimodality Network**
Node01	H_0	8.3%	0.6%
	H_1	**79.0%**	**94.2%**
	H_2	12.7%	5.2%
Node12	H_0	7.6%	6.5%
	H_1	**79.3%**	**88.7%**
	H_2	13.1%	4.8%

6: Conclusion

Detecting radiological materials that could potentially harm the general public or high value assets has remained a high priority in the U.S. national security realm since the Manhattan project, now over 70 years ago. As a subset of these operations, tracking the material for interdiction has been a nontrivial task, where the general concept of operations relies on practitioners to aggressively survey the area around the initial detection location. This work has proposed a methodology to utilize a radiation detection sensor network positioned in fixed locations to aid the end-users in the localizing and interdiction process. Three unique modalities were analyzed to increase the ability in tracking radiation anomalies travelling within a sensor network. As the primary sensor modality for alarming, the data collected from the NaI(Tl) gamma-ray radiation detectors were evaluated to identify key features that could aid the operations. The primary feature extracted from the radiation data was the background-subtracted and filtered anomalous spectrum that was utilized to calculate the cosine similarity at other nodes in the network. This feature provided a key capability (i.e. perhaps *the* key capability) to discriminate between the hypotheses when the anomalous spectra contained spectral features, such as energy peaks.

As the sole contextual sensors evaluated in this work, the video camera and LIDAR sensor data were first analyzed to extract useful features. The video data was analyzed using an open-source, COTS solution (i.e. YOLOv3 neural network) that detects, tracks, and classifies objects within each frame of the video feed. The processing of the LIDAR data required more effort to analyze than originally expected, as there currently exist only early-TRL solutions with sometimes questionable results [109]. The LIDAR data was processed and analyzed by piping the point cloud through the open-source ROS software and importing the point cloud into

Python. Once in the Python environment, the point cloud was down sampled where an object detection and tracking became possible.

The Bayesian-based decision-making architecture created and evaluated in this work contains the ability to digest different data streams and perform hypothesis tests in real-time. The time required to formally calculate the posteriors from the conditional probabilities and priors for each data feature utilized in the decision-making architecture is extremely limited, making it ideal for real-time analysis. The downfall in the proposed architecture with respect to time-to-decision is the architecture in its current form requires each sensor modality to be analyzed in real-time to feed the process. At any rate, the architecture provided a high level of accuracy for the anomalous events analyzed from the ORNL sensor array.

The Bayesian-based architecture is not without faults. Bayesian networks in general suffer from multiple drawbacks, namely:

1. They require expert-level knowledge to create informative priors (i.e. expert systems),
2. Introduction of bias when designing the probabilities and the interaction between the different data streams,
3. Requires unique methods (e.g., MCMC) to calculate distributions when sparse data is collected or when no physical relationship is available to dictate the probability, and
4. Computationally expensive for large number of model variables.

In the work provided in this report, the primary areas of where error exists are in the strong user-bias related to the creation of the priors and conditional probabilities (e.g., likelihood functions). While there was an attempt to reinforce the decisions made to create the probabilities with informed reasoning from lessons learned on the ORNL sensor array data, the network may provide less than ideal results in other settings. To that end, if the background radiation remains

relatively constant at each node, which may not be the case (e.g., radon washout), and only a few objects are in the scene around the detectors at a single moment, the author is confident that this architecture should provide adequate performance.

7: Future Work

This work is far from complete and the reader should note that in its current state, the research presented provides solely a proof-of-concept and should be further evaluated prior to being utilized in a real-world environment. As in the majority of cases when performing research, more data are needed to identify the problems that may not be highlighted in the few examples analyzed in this work. Additionally, more data would provide the ability to better characterize the Bayesian classification process and provide more quantifiable metrics (e.g., ROC curve). A primary reason for only a few example events being presented in this work is due to the sheer size of the data from each event. This is inherent to multimodal data fusion scenarios, but especially in this work, where the data are first saved and require a user to parse months of data to find anomalous events. To that end, there do exist a few areas of work that could enhance the research presented here.

First, the overall decision-making architecture could be further analyzed to incorporate dependencies between the data. In the Bayesian context, assuming each modality is independent from one another is rather naïve (i.e. a play-on-words of the common naïve Bayes methodology) and better understanding the dependencies on each could improve the classification or hypothesis testing process (e.g., similarity in energy spectra and the expected count rate). A future effort should perform a covariance analysis to provided a better understanding of the dependencies.

Second, an additional data fusion technique for decision-making called Dempster-Shafer Theory (i.e. also known as DST, evidence theory, belief networks), should be evaluated for the problem statement analyzed in this work. DST is a relatively common approach to sensor data fusion for decision-making tasks and is built upon mathematical set theory, rather than purely

statistical probability like the Bayesian framework. DST provides a direct ability to calculate uncertainty from the sets constructed rather than incorporating uncertainty in the probabilities in the Bayesian framework, and also allows for independent items of evidence to contribute at a higher level of abstraction to multiple decisions rather than a single decision selection (e.g., MAP in Bayesian context). For interested readers, several references for the reader's review include [180, 181, 182, 183, 184]. Outside of the Dempster's and Shafer's original works, the author suggests the work by Klein in [182] that provides a much more in-depth analysis in the application space of sensor data fusion and highlights some of the nuisance when constructing sensor fusion architectures.

Additionally, combining weak classifiers to improve an overall system classification performance is not a novel idea. In the data science community, these efforts are referred to as boosting that attempt to train the individual weak classifiers and update the classifiers to improve the classification. Popular techniques such as Adaptive boosting (i.e. Adaboosting) and gradient boosting may provide increased system performance when trained on the data collected from the ORNL sensor array. For readers who are interested in learning more about boosting and the two techniques stated previously, the author suggests reviewing [185, 186] and [187, 188], respectively.

A sensitivity study should be performed on the data analyzed in this work to focus on sensor drop-outs (i.e. momentary or permanents loss of a data stream) and moving to a live-time analysis platform. As in most operational environments, the planned, best-case scenarios typically never exist, and the end-user team is required to remain adaptive to maintain operability. Such scenarios as loss of power, loss of communication, or inoperable sensors are not uncommon and further research could highlight the effect of losing each sensor modality at

subsequent nodes on the decision-making process. Furthermore, an end-user could truly benefit from a sensitivity study to quantify the benefit each sensor modality provides for detection and tracking of radiological anomalies. This would provide end-users the ability to select the appropriate sensors to meet the specific operational need and not rely on selecting multiple modalities to provide similar results.

Lastly, a full characterization of the LIDAR object detection algorithm is required prior to using the approach outside of the data analyzed from the ORNL sensor array. Absent of any formal analysis on the classification metrics (e.g., FPR, TPR, TNR, TNR) and selection of the hyperparameters (e.g., occupancy cell size and timing resolution), the algorithm performed adequately and did not diminish the classification process provided in this work. It is of the author's opinion that LIDAR data could provide additional information as well regarding the detected objects. The LIDAR data also returns a reflectivity of the object which may provide the ability to differentiate between objects (e.g., metallic materials on vehicles, brick walls, fabric on a person, etc.). Additionally, the ability to process 3D LIDAR data in real-time is not impossible and has been shown on several automated vehicle demonstrations using application specific integrated circuits (ASICs). An effort that was experimented by the author of this work was to extract the 3D point cloud (column) from the occupancy cell and use that information for detection and classification.

References

1. R. Byrd, J.M. Moss, W.C. Priedhorsky, C.A. Pura, G.W. Richter, K.J. Saeger, W.R. Scarlett, S.C. Scott, and R.L. Wagner, ., Nuclear Detection to Prevent or Defeat Clandestine Nuclear Attack, Los Alamos Manuscript LA-UR -04-0629, submitted to IEEE Sensors Journal, Special Issue on Sensors for the Prevention of Terrorist Acts (2004).
2. Runkle, R. C., Smith, L. E., and Peurrung, A. J. (2009). The Photon Haystack and Emerging Radiation Detection Technology. Journal of Applied Physics, 106(4). https://doi.org/10.1063/1.3207769.
3. Medalia, J. (2011). Dirty Bomb: Technical Background, Attack Prevention and Response, Issues for Congress. Congressional Research Service. CRS Report No. R41890. Retrieved on: December 2019. Retrieved from: https://fas.org/sgp/crs/nuke/R41890.pdf.
4. Nuclear Threat Initiative. Radiological Terrorism and Nuclear Security - Module 3: Radiological Terrorism and Nuclear Sabotage. Retrieved on: December 2019. Retrieved from: https://tutorials.nti.org/nuclear-and-radiological-security/terrorism-radiological/.
5. Nuclear Threat Initiative. (2019). CNS Global Incidents and Trafficking Database. Retrieved on: December 2019. Retrieved from: https://media.nti.org/documents/global_incidents_trafficking_2018.pdf.
6. Brennan, S. M., Mielke, A. M., Torney, D. C., and Maccabe, A. B. (2004). Radiation detection with distributed sensor networks. Computer, 37(8), 5759. https://doi.org/10.1109/MC.2004.103.
7. Mielke, A. M., Brennan, S. M., Smith, M. C., Torney, D. C., Maccabe, A. B., and Karlin, J. F. (2005). Independent sensor net- works. IEEE Instrumentation and Measurement Magazine, 8(2), 3337. https://doi.org/10.1109/MIM.2005.1438842.
8. The Royal Society. (2008). Detecting nuclear and radiological materials Workshop. p.1-33.
9. Liu, A., Wu, M., Chandy, K. M., Obenshain, D., Smith, M., and Mclean, R. (n.d.). (2010) Design Tradeoffs for Radiation Detection Sensor Networks.California Institute of Technology. Pasadena, CA.
10. Schulze, J., M. Auer, and R. Werzi. (2000). Low level radioactivity measurement in support of the CTBTO. Applied Radiation and Isotopes. V52:p.23-30.
11. Hart, J. K., and Martinez, K. (2006). Environmental Sensor Networks: A revolution in the earth system science? Earth-Science Reviews, 78(34), 177191. https://doi.org/10.1016/j.earscirev.2006.05.001.
12. Environmental Protection Agency. (1988). Environmental Radiation Ambient Monitoring System - Project Overview. EPA 520/5-84-007.

13. Tsahi Gozani, ed. (1981). Active Nondestructive Assay of Nuclear Materials, Principles and Applications. U.S. Nuclear Regulatory Commission. DOI: 10.2172/6215952/.

14. Reilly, D., Ensslin, N., and Smith, H. Jr. (1991). Passive nondestructive assay of nuclear materials (NUREG/CR–5550). U.S. Nuclear Regulatory Commission. Retrieved from: https://www.lanl.gov/org/ddste/aldgs/sst-training/technical-references.php.

15. Ferguson, C. and W. Potter. The Four Faces of Nuclear Terrorism. (2004). Center for Nonproliferation Studies. Monterey, CA. ISBN: 978-0415952439.

16. Murray, A., R. Marten, A. Johnston, and P. Martin. Analysis for naturally occurring radionuclides at environmental concentrations by gamma spectrometry. In: Journal of Radioanalytical and Nuclear Chemistry 115.2 (1987), pp. 263–288.

17. Knoll, G. F. Radiation Detection and Measurement. John Wiley and Sons, 2010. ISBN: 9780470131480.

18. Fawcett, Tom (2006). An Introduction to ROC Analysis. Pattern Recognition Letters. 27 (8): 861–874. DOI:10.1016/j.patrec.2005.10.010.

19. Receiver Operator Curve. (2020). *Wikipedia*. Retrieved from: https://en.wikipedia.org/wiki/Receiver_operating_characteristic.

20. Macskassy, S., and F. Provost. (2004). Confidence Bands for ROC Curves: Methods and an Empirical Study. Conference on ROC Analysis in Artificial Intelligence. Valencia, Spain.

21. Wallis, S. (2013). Binomial confidence intervals and contingency tests: mathematical fundamentals and the evaluation of alternative methods. Journal of Quantitative Linguistics. 20 (3): 178–208. DOI:10.1080/09296174.2013.799918.

22. Wilson, E. (1927). Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association. 22 (158): 209–212. DOI:10.1080/01621459.1927.10502953. JSTOR 2276774.

23. Newcombe, R. (1998). "Two-sided confidence intervals for the single proportion: comparison of seven methods". Statistics in Medicine. 17 (8): 857–872. DOI:10.1002/(SICI)1097-0258[.

24. Clopper, C. and E. Pearson. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 26 (4): 404–413. DOI:10.1093/biomet/26.4.404.

25. Thulin, M. (2014-01-01). The cost of using exact confidence intervals for a binomial proportion. Electronic Journal of Statistics. 8 (1): 817–840. arXiv:1303.1288. DOI:10.1214/14-EJS909. ISSN 1935-7524.

26. Brown, L., T. Cai, and A. DasGupta. (2001). Interval Estimation for a Binomial Proportion. Statistical Science. 16 (2): 101–133. CiteSeerX 10.1.1.50.3025. DOI:10.1214/ss/1009213286.

27. Ziegler, J.F. Terrestrial Cosmic Rays. (1996). In: IBM Journal of Research and Development. ISSN: 0018-8646. DOI: 10.1147/rd.401.0019.

28. Ziock, K. P. (2002). The Lost Source, Varying Backgrounds and Why Bigger May Not Be Better. AIP Conference Proceedings, 632(1), 6070. https://doi.org/10.1063/1.1513955.

29. Aucott, T. J., Chivers, D. H., & Vetter, K. (2011). Proximity localization with the Mobile Imaging and Spectroscopic Threat Identification (MISTI) system. In *2011 IEEE Nuclear Science Symposium Conference Record* (pp. 1303–1305). IEEE. https://doi.org/10.1109/NSSMIC.2011.6154330.

30. Willis, M., I. Stewart, A. Nicholson, D. Archer, and W. Ray. Systematic Study of Variation in Radiation Data in Cluttered Environments. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, (2018) ISSN: 0168-9002.

31. Xiong, S. (2014). Detection and Estimation Final Project Report: Modeling and the K-Sigma Algorithm for Radiation Detection. New Brunswick, New Jersey.

32. Iyengar, Anagha, Beach, Matthew, Newby, Robert J., Fabris, Lorenzo, Heilbronn, Lawrence H., and Hayward, Jason P. Systematic measurement of fast neutron background fluctuations in an urban area using a mobile detection system. United States: N. p., 2015. Web. doi:10.1016/j.nima.2014.10.045.

33. Archer, D. E., Johnson, J. O., Miller, T. M., Nicholson, A. D., Patton, B. W., and Peplow, D. E. (2015). Systematic Assessment of Neutron and Gamma Backgrounds Relevant to Operational Modeling and Detection Technology Implementation. Oak Ridge National Laboratory, (ORNL/TM-2014/687).

34. Seeman, B. (Nov 1975). US patent document 3,922,541.

35. Saucke, K., G. Pausch, J. Stein, H.G. Ortlepp, and P. Schotanus. Stabilizing Scintillation Detector Systems with Pulsed LEDs: A Method to Derive the LED Temperature from Pulse Height Spectra. IEEE Transactions on Nuclear Science. 52(6):3160-3165. DOI: 10.1109/TNS.2005.862929.

36. Mills, D. L. Internet time synchronization: the network time protocol. IEEE Transactions on communications. p. 1482 - 1493 (1991). DOI: 10.1109/26.103043.

37. Liu, A., Wu, M., Chandy, K. M., Obenshain, D., Smith, M., and Mclean, R. (n.d.). Design Tradeoffs for Radiation Detection Sensor Networks. Thesis (2010). California Institute of Technology. Pasadena, CA.

38. Rao, N. S. V., Shankar, M., Chin, J.-C., Yau, D. K. Y., Srivath- san, S., Iyengar, S. S., Hou, J. C. (2008). Identification of Low- Level Point Radiation Sources Using a Sensor Network. International Conference on Information Processing in Sensor Networks, 493504. https://doi.org/10.1109/IPSN.2008.19.

39. Nemzek, R. J., Dreicer, J. S., Torney, D. C., and Warnock, T. T. (2004). Distributed sensor networks for detection of mobile radioactive sources. IEEE Transactions on Nuclear Science, 51(4 I), 16931700. https://doi.org/10.1109/TNS.2004.832582.

40. Morelande, M. R., and Ristic, B. (2009). Radiological source detection and localisation using Bayesian techniques. IEEE Transactions on Signal Processing, 57(11), 42204231. https://doi.org/10.1109/TSP.2009.2026618.

41. Liu, A. H., Bunn, J. J., and Chandy, K. M. (2010). An analysis of data fusion for radiation detection and localization. In 2010 13th International Conference on Information Fusion (pp. 18). Edinburgh, UK: IEEE. https://doi.org/10.1109/icif.2010.5711879.

42. Shenton-Taylor, C., Gillespie, S., Blackie, D., Caister, A., Bell, A., Murphy, R., Shute, N. (2016). Distributed detector radiation networks: Static and mobile sensor configurations. 2014 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2014. https://doi.org/10.1109/NSSMIC.2014.7431093.

43. Alagha, A., S. Singh, R. Mizouni, and A. Ouali. Data-Driven Dynamic Active Node Selection for Event Localization in IoT Applications - A Case Study of Radiation Localization. IEEE Access (Volume 7). p. 16168 - 16183 (2019). DOI: 10.1109/ACCESS.2019.2894956.

44. Whipps, G. T., and Moses, R. L. (2015). Distributed Sensing for Quickest Change Detection of Point Radiation Sources. In 18th International Conference on Information Fusion (pp. 2227). Washington, D.C.: IEEE. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/1027422.pdf.

45. King, M. J., Harris, B., Toolin, M., DuBord, R. M., Skowron- ski, V. J., LuSoto, M. A., Shokhirev, K. N. (2010). An urban environment simulation framework for evaluating novel distributed radiation detection architectures. 2010 IEEE International Conference on Technologies for Homeland Security (HST), 01810, 446452. https://doi.org/10.1109/THS.2010.5654958.

46. Stephens, D. L., and Peurrung, A. J. (2004). Detection of moving radioactive sources using sensor networks. IEEE Transactions on Nuclear Science, 51(5), 22732278. https://doi.org/10.1109/TNS.2004.834713.

47. Hite, Jason, Mattingly, John, Archer, Daniel E., Willis, Michael J., Rowe, Andrew J., Bray, Kayleigh B., Carter, Jr., Jake, and Ghawaly, James M. Localization of a radioactive source in an urban environment using Bayesian Metropolis methods. United States: N. p., 2018. Web. doi:10.1016/j.nima.2018.09.032.

48. U.S. Department of Energy. *National Laboratory Network*. (2020) Retrieved from https://www.energy.gov/national-laboratories.

49. Anastasio, M. R. *Establishment of the Department of Homeland Security*. Hearing of the U.S. Senate Energy and Natural Resources Committee. Livermore, CA.

50. Hernandez, J. E., & Valentine, J. (2004). 2002-2003 Engineering Accomplishments : Unconventional Nuclear Weapons Detection. In UCRL-TR-204274. Livermore, CA: Lawrence Livermore National Laboratory.

51. Lawrence Livermore National Laboratory. (2004). Science and Technology Review: Atoms for Peace after 50 Years. Livermore, CA.

52. Lawrence Livermore National Laboratory. (2005). Science and Technology Review: Award-Winning Science and Technology 2005 R&D 100 Awards. Livermore, CA. Available on: https://doi.org/10.1176/pn.37.23.0001c.

53. Cunningham, C. (1995). Detection and Track of Stochastic Target using Multiple Measurements. Livermore, CA.

54. Keillor, M. E. (2000). Principal component analysis of low-resolution energy spectra to identify gamma sources in moving vehicle traffic. Dissertation. Oregon State University.

55. Lanman, D. R. (2001). Distributed Sensor Networks with Collective Computation (Vol. LA-UR-01-4).

56. Brennan, S. M., Mielke, A. M., & Torney, D. C. (2005). Radioactive source detection by sensor networks. IEEE Transactions on Nuclear Science, 52(3), 813–819. https://doi.org/10.1109/TNS.2005.850487.

57. Nemzek, R. J., Dreicer, J. S., Torney, D. C., & Warnock, T. T. (2004). Distributed sensor networks for detection of mobile radioactive sources. IEEE Transactions on Nuclear Science, 51(4 I), 1693–1700. https://doi.org/10.1109/TNS.2004.832582.

58. Lawrence Livermore National Laboratory. (2004). Science and Technology Review: Atoms for Peace after 50 Years. Livermore, CA.

59. U.S. Deparment of Homeland Security. Intelligent Radiation Sensing System. Available on: https://www.dhs.gov/intelligent-radiation-sensing- system. Retrieved on: December 2019.

60. Cooper, D. A., Ledoux, R. J., Kamieniecki, K., Korbly, S. E., Thompson, J., Batcheler, J., Aiyawar, V. (2010). Intelligent Radiation Sensor System (IRSS) advanced technology demonstrator (ATD). 2010 IEEE International Conference on Technologies for Homeland Security, 414420. https://doi.org/10.1109/THS.2012.6459901.

61. Cooper, D. A., Ledoux, R. J., Kamieniecki, K., Korbly, S. E., Thompson, J., Batcheler, J., Aiyawar, V. (2012). Intelligent Radiation Sensor System (IRSS) advanced technology demonstrator (ATD). 2012 IEEE International Conference on Technologies for Homeland Security, HST 2012, 511516. https://doi.org/10.1109/THS.2012.6459901.

62. Cooper, D. A., Ledoux, R. J., Kamieniecki, K., Korbly, S. E., Hemp- stead, D., Janney, L., Monnier, C. (2015). Advanced Algorithm Development for Detection, Tracking , and I Identification of Vehicle Radiation Source es in a Multi-Sensor, Distributed Testbed.

63. U.S. Department of Homeland Security. (2017). Emerging Technologies for Large Event Monitoring. Presentation. Washington, DC.

64. Cooper, D. A. (2017). Performance of the Multi-sensor Interdiction System Testbed (MIST) Advanced Technology Demonstration (ATD) for Detection and Tracking of Vehicle-borne Radiation Sources. In IEEE Nuclear Science Symposium Conference (pp. 1–14). Atlanta, GA: IEEE.

65. DARPA SIGMA. (2014). SIGMA. [online] Darpa.mil. Available at: https://www.darpa.mil/program/sigma.

66. DARPA SIGMA. (2016). SIGMA. [online] Darpa.mil. Available at: https://www.darpa.mil/about-us/timeline/sigma.

67. Kromek Group PLC. D3S NET Networked Radiation Detector. Retrieved from: https://www.kromek.com/product/d3s-net/. Retrieved on: December 2019.

68. DARPA SIGMA+. (2017). Radioactive Threat Detection System Completes Emergency Vehicle Test Deployment in Nation's Capital https://www.darpa.mil/news-events/2017-03-01. Slide(s) 6.

69. U.S. Department of Homeland Security. (2016). Domestic Nuclear Detection Office - Advanced Technology Demonstrations Report. Washington, DC. Available on: https://www.dhs.gov/sites/default/files/publications/ATD%20Fact%20Sheet%20%28April%202016%29_0.pdf.

70. U.S. Department of Homeland Security. (2017). Emerging Technologies for Large Event Monitoring. Presentation. Washington, DC.

71. Curtis, J. C., Cooper, R. J., Joshi, T. H., Cosofret, B., Schmit, T., Wright, J., … Speicher, J. (2018). Simulation and validation of the Mobile Urban Radiation Search (MURS) gamma-ray detector response. Nuclear Instruments and Methods in Physics Research, Section A:

Accelerators, Spectrometers, Detectors and Associated Equipment, 954 (August 2018), 161128. https://doi.org/10.1016/j.nima.2018.08.087.

72. Vetter, K. (2016). Multi-sensor radiation detection, imaging, and fusion. Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 805, 127–134. https://doi.org/10.1016/j.nima.2015.08.078.

73. Bandstra, M. S., Aucott, T. J., Brubaker, E., Chivers, D. H., Cooper, R. J., Curtis, J. C., … Vetter, K. (2016). RadMAP: The Radiological Multi-sensor Analysis Platform. Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 840, 59–68. https://doi.org/10.1016/j.nima.2016.09.040.

74. Bandstra, M. S., Quiter, B. J., Curtis, J. C., Bilton, K. J., Joshi, T. H. Y., Meyer, R., … McLean, M. S. L. (2018). Attribution of gamma-ray background collected by a mobile detector system to its surroundings using panoramic video. Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 954 (August 2018). https://doi.org/10.1016/j.nima.2018.08.085.

75. Turin, G. L. (1960). An Introduction to Matched Filters. IRE Transactions on Information Theory. 6 (3): 311–329. doi:10.1109/TIT.1960.1057571.

76. Henderson, K., Stadnikia, K., Martin, A., Enqvist, A., & Koppal, S. (2017). Tracking Radioactive Sources through Sensor Fusion of Omnidirectional LIDAR and Isotropic Rad-Detectors. In International Conference on 3D Vision (pp. 97–106). Qingdao, China. https://doi.org/10.1109/3DV.2017.00021.

77. Stadnikia, K., Martin, A., Henderson, K., Koppal, S., & Enqvist, A. (2018). Data-Fusion for a Vision-Aided Radiological Detection System: Sensor dependence and Source Tracking. In Advancements in Nuclear Instrumentation Measurement Methods and their Applications (Vol. 170, pp. 1–6). https://doi.org/10.1051/epjconf/201817007013.

78. Stadnikia, K., Henderson, K., Koppal, S., & Enqvist, A. (2018). Data Fusion for a Vision-aided Radiological Detection System: Correlation Methods for Multi-source Scenarios. In IEEE Nuclear Science Symposium and Medical Imaging Conference (pp. 1–4). Sydney, Australia: IEEE. https://doi.org/10.1109/NSSMIC.2018.8824275.

79. Alpha Spectra, Inc. (2020). Scintillation Detectors. Retrieved from: https://alphaspectra.com/.

80. David Instruments. (2019). Vantage Pro2 Weather Station. Retrieved from: https://www.davisinstruments.com/solution/vantage-pro2/.

81. Velodyne Lidar, Inc. (2020). Puck Lidar Sensor, High-Value Surround Lidar. Retrieved from: https://velodynelidar.com/products/puck/.

82. Ubiquiti, Inc. (2020). Ubiquiti - UniFi Video Camera G3. Retrieved from: https://www.ui.com/unifi-video/unifi-video-camera-g3/.

83. Seeman, B. (Nov 1975). US patent document 3,922,541. Available from INIS: http://inis.iaea.org/search/search.aspx?orig_q=RN:08284304.

84. Compagnie de Saint-Gobain S.A. (2020). Gain Stabilization Methods. Retrieved form: https://www.crystals.saint-gobain.com/sites/imdf.crystals.com/files/documents/gain_stabilization_methods.pdf.

85. Bl'azquez-Garc'ia, A., Conde, A., Mori, U., & Lozano, J.A. (2020). A review on outlier/anomaly detection in time series data. ArXiv, abs/2002.04236.

86. Hodge, V. J., & Austin, J. (2004). A Survey of Outlier Detection Methodologies. Artificial Intelligence Review, 22(1969), 85–126. https://doi.org/10.1007/s10462-004-4304-y.

87. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection - A Survey. ACM Computing Surveys, 41(3), 1–58. https://doi.org/10.1145/1541880.1541882.

88. Pfund, D. M., Anderson, K. K., Detwiler, R. S., Jarman, K. D., McDonald, B. S., Milbrath, B. D., … Woodring, M. L. (2016). Improvements in the method of radiation anomaly detection by spectral comparison ratios. Applied Radiation and Isotopes, 110, 174–182. https://doi.org/10.1016/j.apradiso.2015.12.063.

89. Anderson, K. K., Jarman, K. D., Mann, M. L., Pfund, D. M., & Runkle, R. C. (2008). Discriminating nuclear threats from benign sources in gamma-ray spectra using a spectral comparison ratio method. Journal of Radioanalytical and Nuclear Chemistry, 276(3), 713–718. https://doi.org/10.1007/s10967-008-0622-x.

90. Jarman, K. D., Smith, L. E., & Carlson, D. K. (2004). Sequential probability ratio test for long-term radiation monitoring. IEEE Transactions on Nuclear Science, 51(4 I), 1662–1666. https://doi.org/10.1109/TNS.2004.832543.

91. Liu, A. H., Julian, B. J., & Mani, C. K. (2010). An Analysis of Data Fusion For Radiation Detection and Localization. In 13th Conference on Information Fusion. Edinburgh, UK: IEEE. https://doi.org/10.1109/ICIF.2010.5711879.

92. Kar, S. and A. Ramalingam. (2013). Is 30 the Magic Number? Issues in Sample Size. National Journal of Community Medicine. Volume 4:1. pISSN 0976-3325.

93. Chakrapani, C. (2011). Statistical Reasoning vs. Magical Thinking. VUE, Magazine of the MRIA. Retrieved from: http://www.chuckchakrapani.com/Articles/PDF/0411Chakrapani.pdf.

94. Kadam, P. and S. Bhalerao. Sample size calculation. International journal of Ayurveda research vol. 1,1 (2010): 55-7. DOI:10.4103/0974-7788.59946.

95. Montgomery, D. and C. Runger. (2014). Applied Statistics and Probability for Engineers (6th ed.). Wiley. p.241. ISBN 9781118539712.

96. Aucott, T. (2014). Gamma-Ray Background Variability in Mobile Detectors. PhD Dissertation. University of California, Berkeley. https://doi.org/10.1300/J122v22n03_06.

97. National Oceanic and Atmospheric Administration (NOAA). (2013). LIDAR — Light Detection and Ranging — is a remote sensing method used to examine the surface of the Earth. Retrieved from: http://oceanservice.noaa.gov/facts/lidar.html.

98. J. Shan and C. Toth. (2018).Topographic Laser Ranging and Scanning: Principles and Processing, 2nd Ed. CRC Press.

99. Cracknell, A. and L. Hayes. (2007). Introduction to Remote Sensing, 2nd Ed. London: Taylor and Francis. ISBN 978-0-8493-9255-9.

100. Lim, H. and A. Taeihagh. (2019). Algorithmic Decision-Making in AVs: Understanding Ethical and Technical Concerns for Smart Cities. Sustainability. 11 (20): 5791. DOI:10.3390/su11205791.

101. Royo, S. and M. Ballesta-Garcia. (2019). An Overview of Lidar Imaging Systems for Autonomous Vehicles. Applied Sciences, 9(19), 4093. Retrieved from: https://doi.org/10.3390/app9194093.

102. Weitkamp, C. (2005). LiDAR: Introduction. In Laser Remote Sensing; CRC Press: Boca Raton, FL, USA; pp. 19–54.

103. Comerón, A. and C. Munoz-Porcar, F. Rocadenbosch, A. Rodríguez-Gómez, Sicard, M. (2017). Current research in LIDAR technology used for the remote sensing of atmospheric aerosols. Sensors 17:1450.

104. Dong, P. and Q. Chen. (2018). LiDAR Remote Sensing and Applications; CRC Press: Boca Raton, FL, USA.

105. Velodyne Lidar, Inc. (2016). User's Manual and Programming Guide: VLP-16 Velodyne LiDAR Puck.

106. Businesswire. (2018). Velodyne Lidar Announces Collaboration with Nikon in Technology Development and Manufacturing. Retrieved from: https://www.businesswire.com/news/home/20181220006016/en/Velodyne-Lidar-Announces-Collaboration-Nikon-Technology-Development.

107. U.S. Congress. (2015). LAS (LASer) File Format, Version 1.4. Library of Congress. Retrieved from: https://www.loc.gov/preservation/digital/formats/fdd/fdd000418.shtml.

108. The American Society for Photogrammetry & Remote Sensing. (2013). LAS Specification Verstion 1.4 - R13. Retrieved from: https://www.asprs.org/wp-content/uploads/2010/12/LAS_1_4_r13.pdf.

109. Stanford Artificial Intelligence Laboratory et al. (2018). Robotic Operating System. Retrieved from https://www.ros.org.

110. Meng, X., N. Currit, and K. Zhao. (2010). Ground Filtering Algorithms for Airborne LiDAR Data: A Review of Critical Issues. Remote Sensing, 2(3), 833-860. ISSN: 2072-4292. Retrieved from: https://doi.org/10.3390/rs2030833.

111. Vu, Hoang, H. Tong, P. Chu, W. Zhang, S. Cho, Y. Park, and K. Cho. (2017). Adaptive ground segmentation method for real-time mobile robot control. International Journal of Advanced Robotic Systems, 14(6). Retrieved from: https://doi.org/10.1177/1729881417748135.

112. Thrun, S., W. Burgard, and D. Fox. (2005). Probabilistic Robotics, MIT Press. Retrieved from: http://www.probabilistic-robotics.org/.

113. C. Chen, C. Tay, C. Laugier and K. Mekhnacha. (2006). Dynamic environment modeling with gridmap: A multiple-object tracking application. International Conference on Control Automation Robotics and Vision.

114. Kalman, R. E. (1960). “A New Approach to Linear Filtering and Prediction Problems,” Transaction of the ASME—Journal of Basic Engineering, pp. 35-45.

115. Sorenson, H. W. (1970). “Least-Squares estimation: from Gauss to Kalman,” IEEE Spectrum, vol. 7, pp. 63-68, July 1970.

116. Welch, G. and G. Bishop. (1995) "An Introduction to Kalman Filter". Department of Computer Science. University of North Carolina. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.336.5576&rep=rep1&type=pdf.

117. Chui, C. and G. Chen. (2017). "Kalman Filter with Real-time Applications". Springer International Publishing. Retrieved from: https://doi.org/10.1007/978-3-319-47612-4.

118. Oliphant, T. E. (2006). A guide to NumPy (Vol. 1). Trelgol Publishing USA.

119. Pappas, T. (2011). New Challenges for Image Processing Research. IEEE Transactions on Image Processing, 20(12). DOI:10.1109/TIP.2011.2172110.

120. Kang, B. (2007). A Review on Image and Video Processing. International Journal of Multimedia and Ubiquitous Engineering, 2(2).

121. Hadi, R., G. Sulong, and L. George. (2014). Vehicle Detection and Tracking Techniques: A Concise Review. Signal & Processing: An International Journal, 5(1). Retrieved from: https://arxiv.org/pdf/1410.5894.pdf.

122. Jalled, F. and I. Voronkov. (2016). Object Detection using Image Processing. arXiv Archive, Computer Science - Computer Vision and Patter Recognition. arXiv No: 1611.07791v1. Retrieved from: https://arxiv.org/abs/1611.07791.

123. Kulchandani, J. and K. Dangarwala. (2015). Moving Object Detection: Review of Recent Research Trends. 2015 International Conference on Pervasive Computing.

124. Egmont-Peterson, M., D. de Ridder, and H. Handels. (2002). Image Processing with Neural Networks - a Review. Pattern Recognition, 35(20), 2279-2301. Retrieved from: https://doi.org/10.1016/S0031-3203(01)00178-9. DOI: 10.1109/PERVASIVE.2015.7087138.

125. Szeliski, R. (2010). Computer Vision: Algorithms and Applications, 1st Edition.

126. Prince, S. (2012). Computer Vision: Models, Learning, and Inference. Cambridge University Press. ISBN: 978-1107011793.

127. Forsyth, D. and J. Ponce. (2011). Computer Vision: A Modern Approach. Pearson Publishing. ISBN: 978-0136085928.

128. Tang, J. (2018). Intelligent Mobile Projects with TensorFlow. Packt Publishing. ISBN: 9781788834544.

129. InceptionV3. (2017). RSNA Pneumonia Detection Challenge - Kaggle. Additional information available at: https://www.kaggle.com/keras/inceptionv3/home.

130. Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Zbigniew Wojna. (2015). Rethinking the Inception Architecture for Computer Vision. arXiv Archive, Computer Science - Computer Vision and Patter Recognition. Retrieved from: https://arxiv.org/abs/1512.00567.

131. Lin, T., P. Goyal, R. Girshick, K. He, and P. Dollar. (2018). Focal Loss for Dense Object Detection. arXiv Archive, Computer Science - Computer Vision and Patter Recognition. Retrieved from: https://arxiv.org/pdf/1708.02002.pdf.

132. Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg. (2015). SSD: Single Shot Multibox Detector. arXiv Archive, Computer Science - Computer Vision and Patter Recognition. Retrieved from: https://arxiv.org/abs/1512.02325.

133. Fu, C., W. Liu, A. Ranga, A. Tyagi, and A. Berg. (2017). DSSD : Deconvolutional single shot detector. arXiv Archive, Computer Science - Computer Vision and Patter Recognition. Retrieved from: https://arxiv.org/abs/1701.06659.

134. Girshick, R. B. (2015). Fast R-CNN. arXiv Archive, Computer Science - Computer Vision and Patter Recognition. Retrieved from: https://arxiv.org/abs/1504.08083.

135. Girshick, R. B., Donahue, J., Darrell, T., and Malik, J. (2013). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. arXiv Archive, Computer Science - Computer Vision and Patter Recognition. Retrieved from: https://arxiv.org/abs/1311.2524.

136. Ren, S., K. He, R. Girshick, and J. Sun. (2015). Faster R-CNN: Toward Real-time Object Detection with Region Proposal Networks. arXiv Archive, Computer Science - Computer Vision and Patter Recognition. Retrieved from: https://arxiv.org/abs/1506.01497.

137. Redmon, J., S. Divvala, R. Girshick, and A. Farhadi. (2015). You only look once: Unified, real-time object detection. arXiv Archive, Computer Science - Computer Vision and Patter Recognition. arXiv:1506.02640v5. Retrieved from: https://arxiv.org/abs/1506.02640.

138. Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv Archive, Computer Science - Computer Vision and Patter Recognition. Retrieved from: https://arxiv.org/abs/1804.02767.

139. Alpaydin, E. (2018). Introduction to Machine Learning. 3rd Edition. MIT Press, Cambridge, MA, USA. ISBN: 978-81-203-5078-6.

140. Dixit, K., M. Chadaga, S. Salalgimath, G. Rakshith, and N. Kumar. (2019). Evaluation and Evolution of Object Detection Techniques YOLO and R-CNN. International Journal of Recent Technology and Engineering. ISSN: 2277-3878, Volume-8, Issue-2S3. Retrieved from: https://www.ijrte.org/wp-content/uploads/papers/v8i2S3/B11540782S319.pdf.

141. Pulli, K., A. Baksheev, K. Kornyakov, and V. Eruhimov. (2012). Real-time computer vision with OpenCV, Communications of the ACM. DOI:10.1145/2184319.2184337.

142. Abadi, M., et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org. https://www.tensorflow.org/.

143. Zhao, Z., P. Zheng, S. Xu, and X. Wu. (2019). Object Detection with Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems. Retrieved from: https://arxiv.org/abs/1807.05511.

144. Redmon, J., and A. Farhadi. (2016). YOLO9000: Better, Faster, Stronger. arXiv Archive, Computer Science - Computer Vision and Patter Recognition. Retrieved from: https://arxiv.org/abs/1612.08242.

145. Lin, T., et al. (2014). Microsoft COCO: Common Objects in Context. arXiv Archive, Computer Science - Computer Vision and Patter Recognition. Retrieved from: https://arxiv.org/abs/1405.0312.

146. Tandon, P., Huggins, P., Maclachlan, R., Dubrawski, A., Nelson, K., & Labov, S. (2016). Detection of radioactive sources in urban scenes using Bayesian Aggregation of data from

mobile spectrometers. Information Systems, 57, 195–206. https://doi.org/10.1016/j.is.2015.10.006

147. Tandon, P. (2016). Bayesian aggregation of evidence for detection and characterization of patterns in multiple noisy observations. Carnegie Mellon University.

148. Gosnell, T. B., Hall, J. M., Ham, C. L., Knapp, D. A., Koenig, Z. M., Luke, S. J., … Wolford, J. K. (1997). Gamma-Ray Identification of Nuclear Weapon Materials. Livermore, CA. https://doi.org/UCRL-ID-127436.

149. Fagan, D. K., Robinson, S. M., & Runkle, R. C. (2012). Statistical methods applied to gamma-ray spectroscopy algorithms in nuclear security missions. *Applied Radiation and Isotopes*, *70*(10), 2428–2439. https://doi.org/10.1016/j.apradiso.2012.06.016.

150. Los Alamos Scientific Laboratory. (1979). Group X-6. MCNP : a General Monte Carlo Code for Neutron and Photon Transport. Los Alamos, N.M. : [Springfield, Va.] :Dept. of Energy, Los Alamos Scientific Laboratory.

151. Tammero, L. B., Lange, D., & Santiago, C. (2015). Implementation and Field Demonstration of Supervisory Logic Algorithms for Enhanced Radiation Detection. Livermore, CA.

152. K. Saucke, G. Pausch, J. Stein, H. Ortlepp and P. Schotanus, "Stabilizing scintillation detector systems with pulsed LEDs: a method to derive the LED temperature from pulse height spectra," in IEEE Transactions on Nuclear Science, vol. 52, no. 6, pp. 3160-3165, Dec. 2005, doi: 10.1109/TNS.2005.862929.

153. Sadeghi, M., & Behnia, F. (2018). Optimum window length of Savitzky-Golay filters with arbitrary order. Retrieved from http://arxiv.org/abs/1808.10489.

154. Benhamou, E. and V. Melot. (2018). Seven proofs of the Pearson Chi-squared independence test and its graphical interpretation. arXiv:1808.09171. Retrieved from: https://arxiv.org/pdf/1808.09171.pdf.

155. Lopes R.H.C. (2011) Kolmogorov-Smirnov Test. In: Lovric M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg. Retrieved from: https://doi.org/10.1007/978-3-642-04898-2_326.

156. Alexopoulos E. C. (2010). Introduction to multivariate regression analysis. Hippokratia, 14(Suppl 1), 23–28. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3049417/.

157. Tandon, P. (2015). Bayesian aggregation of evidence for detection and characterization of patterns in multiple noisy observations. Carnegie Mellon University. Retrieved from http://dl.acm.org/citation.cfm?doid=2911172.2911181.

158. Šidák, Z.K. Rectangular Confidence Regions for the Means of Multivariate Normal Distribution. Journal of the American Statistical Association 62 (318): 626–633; 1967.

159. Hoerl, Arthur E. (1962). "Application of Ridge Analysis to Regression Problems". Chemical Engineering Progress. 58 (3): 54–59.

160. Hoerl, A. E., & Kennard, R. W. (1976). Ridge Regression Iterative Estimation of the Biasing Parameter. Communications in Statistics - Theory and Methods, 5(1), 77–88. https://doi.org/10.1080/03610927608827333.

161. Heinze G, Wallisch C, Dunkler D. (2018). Variable selection - A review and recommendations for the practicing statistician. Biom J.;60(3):431-449. doi:10.1002/bimj.201700067

162. Atkinson, A. C. (1980). A note on the generalized information criterion for choice of a model. Biometrika, 67(2), 413–418. https://doi.org/10.1093/biomet/67.2.413.

163. Khalaf, G., & Shukur, G. (2005). Choosing Ridge Parameter for Regression Problems. Communications in Statistics - Theory and Methods, 34(5), 1177–1182. https://doi.org/10.1081/STA-200056836.

164. Liu, S., & Dobriban, E. (2019). Ridge Regression: Structure, Cross-Validation, and Sketching, 1–27. Retrieved from http://arxiv.org/abs/1910.02373

165. Presti, C., D. Weier, R. Kouzes, and J. Schweppe. (2006). Baseline suppression of vehicle portal monitor gamma count profiles: A characterization study. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 562:281-297. 10.1016/j.nima.2006.02.156.

166. Coulon, R. (2018). Recent Developments in Count Rate Processing Associated with Radiation Monitoring Systems. 10.5772/intechopen.71233.

167. Siciliano, Edward R., et al. Comparison of PVT and NaI (Tl) scintillators for vehicle portal monitor applications. (2005). Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 550.3: 647-674.

168. Stromswold, David C., et al. "Field tests of a NaI (Tl)-based vehicle portal monitor at border crossings." IEEE Symposium Conference Record Nuclear Science 2004.. Vol. 1. IEEE, 2004.

169. Stewart, I.R., Nicholson, A.D., Archer, D.E. et al. Understanding and quantifying the systematic effects of clutter within a radiation detection scene. J Radioanal Nucl Chem 318, 727–737 (2018). https://doi.org/10.1007/s10967-018-6159-8

170. Willis, M. J., Stewart, I. R., Nicholson, A. D., Archer, D. E., & Ray, W. R. (2020). Systematic study of variation in radiation data in cluttered environments. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 954, 161290. doi:10.1016/j.nima.2018.09.125.

171. Duda, R. O., Hart, P. E., & Stork, D. G. (2000). Pattern Classification. (2nd ed.). Wiley-Interscience. Retrieved from http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471056693.html.

172. Tang, Z., Naphade, M., Liu, M., Yang, X., Birchfield, S., Wang, S., . . . Hwang, J. (2019). CityFlow: A City-Scale Benchmark for Multi-Target Multi-Camera Vehicle Tracking and Re-Identification. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2019.00900.

173. Zheng Tang, Gaoang Wang, Hao Xiao, Aotian Zheng, Jenq-Neng Hwang; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2018, pp. 108-115.

174. Yue Yao and Liang Zheng and Xiaodong Yang and Milind Naphade and Tom Gedeon. (2020). Simulating Content Consistent Vehicle Datasets with Attribute Descent. Computer Vision and Pattern Recognition. Retrieved from: https://arxiv.org/abs/1912.08855.

175. Bishop, C.M. (2006). "Pattern Recognition and Machine Learning." 1st Ed. Springer Publishing. ISBN: 978-1493938438.

176. Ott, R. L. and M. Longnecker. (2015). "An Introduction to Statistical Methods and Data Analysis." 7th Ed. Brooks Cole Publishing. ISBN: 9781305269477.

177. Olivier, C., Schölkopf, B., & Zien, A. (2006). Semi-Supervised Learning. Interdisciplinary sciences computational life sciences (Vol. 1). https://doi.org/10.1007/s12539-009-0016-2.

178. Ai, Y, M. Peng, and K. Zhang. (2018). "Edge computing technologies for Internet of Things: a primer." Digital Communications and Networks 4.2: 77-86.

179. Ahmed, E., and M. H. Rehmani. (2017). "Mobile edge computing: opportunities, solutions, and challenges." *Future Generation Computer Systems*, Volume 70: 59-63.

180. Dempster AP. Upper and lower probabilities induced by a muilti-valued mapping. Ann Math Stat 1967;38:325±39.

181. Shafer G. (1976). A mathematical theory of evidence. Princeton: Princeton University Press.

182. Klein, L. A. (1993). A Boolean Algebra Approach to Multiple Sensor Voting Fusion. IEEE Transactions on Aerospace and Electronic Systems, 29(2), 317–327. https://doi.org/10.1109/7.210070.

183. Beynon, M., Curry, B., & Morgan, P. (2000). The Dempster Shafer theory of evidence : an alternative approach to multicriteria decision modelling. The International Journal of Management Science, 28, 37–50.

184. Khan, M. N., & Anwar, S. (2019). Paradox elimination in dempster–shafer combination rule with novel entropy function: Application in decision-level multi-sensor fusion. Sensors (Switzerland), 19(21), 1–14. https://doi.org/10.3390/s19214810.

185. Freund, Yoav, Robert Schapire, and Naoki Abe. "A short introduction to boosting." Journal-Japanese Society For Artificial Intelligence 14.771-780 (1999): 1612.

186. Schapire, R. E., & Freund, Y. (2013). Boosting: Foundations and algorithms. Kybernetes.

187. Hastie, Trevor, et al. (2009). "Multi-class adaboost." Statistics and its Interface 2.3: 349-360.

188. Friedman, J. H. (2002). Stochastic gradient boosting. Computational statistics & data analysis, 38(4), 367-378.

Appendix

The information in this section provides the reader with additional details on concepts discussed in this work.

YOLOv3 Class List

The object detection and recognition algorithm utilized in this work was the YOLOv3 (i.e. You Only Look Once, Version 3). A trained version of the neural network on the COCO dataset is utilized in this work, resulting in 80 unique classes. The full list for the YOLOv3 algorithm used in this work is provided in the Figure 117. Recall that in this work, the majority of the pretrained classifications were deemed not useful and only three remained; that is, person, car, and truck. There are additional classifications that could be deemed useful, such as Bicycle, Bus, or Motorbike as they could both be potentially utilized to transport radioactive materials in nefarious scenarios.

Derivation of Bayes' Theorem

Bayes' Theorem, also commonly referred to as Bayes' Rule, reorganizes the joint probability definition between two events to create a method to, among other things, update a belief function. A joint probability provides the probability that two events by multiplying the probability of one event occurring by the conditional probability that given that event occurring the second event occurs. The conditional probability is another way of expressing the probabilistic causality for a given relationship. This joint probability for Event A and Event B is expressed given the following relationship, $P(A \cap B) = P(A)\ P(B|A)$. Additionally, the joint probability is also equal to the opposite event sequence, $P(A \cap B) = P(B)\ P(A|B)$. Equating the

1	Person	**16**	Cat	**31**	Skis	**46**	Bowl	**61**	Diningtable	**76**	Vase
2	Bicycle	**17**	Dog	**32**	Snowboard	**47**	Banana	**62**	Toilet	**77**	Scissors
3	Car	**18**	Horse	**33**	Sports Ball	**48**	Apple	**63**	Tvmonitor	**78**	Teddy Bear
4	Motorbike	**19**	Sheep	**34**	Kite	**49**	Sandwich	**64**	Laptop	**79**	Hair Drier
5	Aeroplane	**20**	Cow	**35**	Baseball Bat	**50**	Orange	**65**	Mouse	**80**	Toothbrush
6	Bus	**21**	Elephant	**36**	Baseball Glove	**51**	Broccoli	**66**	Remote		
7	Train	**22**	Bear	**37**	Skateboard	**52**	Carrot	**67**	Keyboard		
8	Truck	**23**	Zebra	**38**	Surfboard	**53**	Hot Dog	**68**	Cell Phone		
9	Boat	**24**	Giraffe	**39**	Tennis Racket	**54**	Pizza	**69**	Microwave		
10	Traffic Light	**25**	Backpack	**40**	Bottle	**55**	Donut	**70**	Oven		
11	Fire Hydrant	**26**	Umbrella	**41**	Wine Glass	**56**	Cake	**71**	Toaster		
12	Stop Sign	**27**	Handbag	**42**	Cup	**57**	Chair	**72**	Sink		
13	Parking Meter	**28**	Tie	**43**	Fork	**58**	Sofa	**73**	Refrigerator		
14	Bench	**29**	Suitcase	**44**	Knife	**59**	Pottedplant	**74**	Book		
15	Bird	**30**	Frisbee	**45**	Spoon	**60**	Bed	**75**	Clock		

Figure 117: Class list for YOLOv3.

two relationships provides a method to update the belief for an event based on the relationships previously derived, $P(A)\,P(B|A) = P(B)\,P(A|B)$ and, thus, $P(A|B) = \frac{P(A)\,P(B|A)}{P(B)}$. This final equation is conventional formula known as Bayes' Theorem and is the basis for statistical inference. In addition, portions of the derived relationship are commonly expressed with unique titles, such as the prior probability, posterior probability, and the normalization probability, as provided in the following equation, $posterior = \frac{prior \times likelihood}{evidence}$. In this case, the prior probability, or often referred to as *a priori* information, corresponds to $P(A)$ and illustrates the prior belief that Event A occurs. The posterior is the probability that was derived, $P(A|B)$, and expresses the new belief of the system. The probability $P(B)$ divides the product in the numerator to normalize the probability, resulting in values between the 0.0 to 1.0. In many applications, the normalization probability can be difficult, if not near-impossible, to calculate analytically. Thus, the normalization probability can be removed from the final formula resulting in values outside the original 0.0 to 1.0 range. In such scenarios, the hypothesis tests select the highest value probability.

Vita

Ian received his Bachelor of Science and Master of Science in Nuclear Engineering from the University of Tennessee in Knoxville, Tennessee in 2013 and 2015, respectively. During his tenure in graduate school, Ian focused on applications relating to proliferation detection and nuclear security as a graduate research student at the Institute for Nuclear Security as well as multiple work experiences at Department of Energy national laboratories. After completing a fellowship at the National Nuclear Security Administration in Washington, D.C., Ian returned to the University of Tennessee to finish his Ph.D. while working at U.S. Department of Energy's Oak Ridge National Laboratory. After graduation, Ian plans to continue to focus on national security issues.