DISSERTATION

DEVELOPMENT OF A BAYESIAN LINEAR REGRESSION MODEL FOR THE DETECTION OF A WEAK RADIOLOGICAL SOURCE FROM GAMMA SPECTRA MEASUREMENTS

Submitted by

Matthew Meengs

Department of Environmental and Radiological Health Sciences

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2021

Doctoral Committee:

Advisor: Alexander Brandl

Thomas E. Johnson
Ralf Sudowe
Piotr Kokoszka

ABSTRACT

DEVELOPMENT OF A BAYESIAN LINEAR REGRESSION MODEL FOR THE DETECTION OF A WEAK RADIOLOGICAL SOURCE FROM GAMMA SPECTRA MEASUREMENTS.

The detection of radiation requires the use of statistical tools due to the probabilistic nature of the emission and the interaction properties of radiation. Frequentist statistical methods are traditionally employed towards this end – the most common being the “traditional” method which calculates a decision threshold above which a source is determined to be present. The decision threshold is calculated from a predetermined false positive rate (typically 5%) and is used as a decision parameter. The decision parameter is a statistical tool by which it is determined whether or not a source other than background is present. In radiological conditions where a source is both improbable and weak, and where counting time is limited, the detection of a source becomes progressively more challenging using this traditional method. The detection of clandestine fissile materials presents such a challenge, and with the increasing risk of nuclear proliferation, there exists a growing motivation to research more optimal methods of detection, especially where a missed detection is of such high consequence.

Previous research has been conducted on using a Bayesian model to develop a decision parameter for weak source detection. The use of a Bayesian model has been shown in laboratory settings to outperform the traditional frequentist method. However, the model tested was designed for gross counts only. In the present study, a Bayesian algorithm is being developed and tested that uses the entirety of the gamma spectrum. Specifically, several Bayesian linear regressions are developed and tested which compared different energy ranges in the spectrum. The parameters generated from these linear regressions are tested for their efficacy as decision parameters. With the additional information presented from the entire

spectrum, it is theoretically possible that even further improvements in the detection of a weak source can be achieved.

The results of this research have shown that regressor coefficients via a Bayesian method are effective as decision parameters. The best results, however, were shown only to match the efficacy of the more traditional, frequentist method of detection.

ACKNOWLEDGMENTS

The author of this dissertation was supported in part by three separate U.S. government agencies: the U.S. Department of Homeland Security (DHS) under Grant Award Number, 2014-DN-077-ARI091-01, the Nuclear Regulatory Commission (NRC), and the Mountains and Plains Educational and Research Center (MAP ERC). The contents expressed herein, however, are solely the responsibility of the author and do not necessarily express the views of the DHS, NRC, or the MAP ERC.

I would also like to express my gratitude towards Dr. Thomas Johnson and Dr. Alex Brandl. Their mentorship has been quite valuable throughout these five years at Colorado State University. I look forward to continued partnership.

I would like to give a shout-out to Rion Marcinko, who absolutely insisted that I get my PhD. After completing my master's degree from CSU, I was on the fence regarding further education. Absent Rion's stubborn admonishments, I may never have pursued the PhD.

Lastly, I would like to thank my parents Rich and Leona Meengs, who encouraged me to continue with my PhD during times of frustration and disappointment. Without their support, this dissertation might never have obtained to completion.

TABLE OF CONTENTS

ABSTRACT……..……....ii

ACKNOWLEDGEMENTS…….....iv

INTRODUCTION…….……...1

Motivation for the present study…….……1

Prior use of frequentist statistical tools…….……..2

Prior use of Bayesian statistical tools……7

Frequentist and Bayesian statistics…….……...7

Research scenario vs. field scenario…….……...11

MATERIALS AND METHODS……..……14

Data acquisition and analysis…….……...14

Physics review – Spectroscopy and a brief review of gamma interactions……..……..…15

Procedure…..……17

Assumptions…….……20

Traditional method of radiation detection…..…….……22

ROC curves as the metric for determining optimal detection…….……..23

SPECTRAL OPTIMIZATION ALGORITHM…….....27

Background…….……27

Theory I – Obtaining a metric for bin optimization...……..……...29

Theory II – Towards finding the optimal combination of energy bins in a spectrum……...30

Development – Mathematical construction…….....……..32

Evaluation and additional ideas…….……34

PRESENTATION OF THE BAYESIAN MODELS……..……37

Background…….……37

Introduction…….……..39

Methods I – Full conditionals…….41

Methods II – Monte Carlo simulation……..…42

Methods III – Diagnostics……...…44

Methods IV – The Algorithm…….…48

Methods V – Evaluation.....……51

Spectra…….…52

Results – Testing the eighteen parameters as decision parameters ….……..…….52

Discussion…….....70

Training Background, minimizing priors, and establishing β^* from background……...72

CONCLUSION……..……76

REFERENCES……..……79

APPENDIX A……...…..…82

APPENDIX B……..…84

GLOSSARY……...…92

INTRODUCTION

Radioactive decay is the release of energy from an atomic nucleus as it transitions from a higher energy state to a lower energy state. Although the conservation of energy is assured, the actual physical pathways for these transitions are greatly varied. Different physical pathways will emit different particles at different energies. In this research, only gamma radiation on the order of kilo-electron volts (keV) are considered, (specifically, the energy ranges from about 40 – 1700 keV). The particle and energy of the emission are not governed by classical physics. This implies that not only the mechanism of decay, but also the time of decay and/or the interval between decays cannot be pre-determined. Quantum theory is the most, and in fact the only, theory to successfully model radiation phenomena. The central caveat is that radioactive decay is fundamentally stochastic, and thus only the probability of decay can be predicted.

Motivation for the present study

There is often a strong desire to detect the presence of an unknown radioactive source. This is inherently difficult for two reasons: (1) the ubiquitous presence of background radiation, and (2) the stochastic nature of the phenomenon. Moreover, these uncertainties are in addition to the inevitable systematic and random uncertainties that arise in any scientific measurement (Bevington, Robinson 2003). Consequently, a signal from a source may not be distinguishable from the background signals if the measurable quantities of these radiations are indistinct. Yet in field applications, one must often make an unequivocal decision as to whether a source is present despite the fundamentally probabilistic nature of these processes. Therefore, in determining the presence of a source, statistical tools are required.

In many circumstances, this task can be completed with little difficulty. For instance, it is easy to distinguish a strong or even moderate source from a weak background. Additionally,

with ample time, even a weak source against a comparable background can eventually be distinguished according to the following proportionality

$$(\text{ability to distinguish source form background}) \propto \sqrt{\text{time}}.$$

Unfortunately, these conveniences are not always present. Field applications exist where the suspected source is weak and yet sufficient time for a clean detection is absent. In some instances, a missed detection can result in dire consequences. The smuggling of illicit fissile materials presents to us an obvious example – the signal is low, there is little time, and the consequences of a missed detection are very high. Hence, there exists a motivation to research more optimal statistical methods for the detection of a weak radiological source. Moreover, information provided from the spectrum will be used to accomplish this end. This motivation provides the impetus for the following hypothesis: using spectral information, there exist Bayesian linear regression coefficients which, used as decision parameters, will surpass the efficacy of the traditional frequentist method of detection.

Prior use of frequentist statistical tools

Historically, frequentist statistics has been used for radiation detection. Frequentist statistics is rooted in inductive reasoning – that the characteristics of a population can be inferred through a subset of that population. Population inference relies on imaginary repeated sampling; it is assumed that with random sampling, the sample will be a representative of the population. This can be expressed as follows:

$$(\text{parent parameter}) = \lim_{n\to\infty} (\text{experimental parameter})$$

where *n* is the sample size (Bevington, Robinson 2003). Usually, a statistic is used as an estimator for a parameter. Thus, the above equation can be expressed as

$$\beta = \lim_{n\to\infty} \hat{\beta}.$$

The Neyman-Pearson lemma provides a criterion by which the ability to distinguish between multiple hypotheses (the null and at least one alternative hypothesis) is optimized (Neyman, Pearson 1933). This historic paper provides the basis for hypothesis testing. The lemma deals only with simple hypotheses (e.g., $\mu = 0$). Two hypotheses are typically posited – traditionally called the null hypothesis and the alternative hypothesis. A rejection region for the null hypothesis is calculated to test these hypotheses. The rejection region is the set of all possible outcomes of the data that would lead one to reject the null hypothesis. The mathematical result of this lemma is that the existence of the optimal rejection region exists (i.e., there exists no other rejection region which leads to a test with more statistical power). The test takes the following general form:

$$p(\text{reject } \mathrm{H}_0 | \mathrm{H}_0 \text{is true}) = \alpha$$

where α is the probability of a type I error (i.e., false positive rate). This probability is typically predetermined. In radioactive source detection, the two hypotheses are defined as:

H_0: No radioactive source other than background is present

H_A: A radioactive source other than background is present.

All subsequent frequentist tools that are used for source detection follow the hypothesis testing framework.

Lloyd Currie used the concept of hypothesis testing to standardize how signal detection is performed in radiation physics. He defines three levels of detection: (1) L_c – the signal level above which a signal other than background may be reliably recognized as detected (i.e., something other than background may exist), (2) L_D – the signal level above which one may reliably conclude (with a certain predetermined probability) that a source of radiation other than background is present, and (3) L_Q – the level above which the measurement precision will be reliable for quantitative determinations. The first and lowest of these levels, L_C – called the critical level, is that level (or limit) above which something other than background is said to be

present within a sample (Currie 1968). Conversely, if a measurement is below the critical level, the sample is considered background. One of the features of the critical level (sometimes called the alarm level) is that it prompts further investigation.

The L_C derived by Currie assumes a Gaussian distribution. Mathematically, the critical level is given by

$$L_C = k_\alpha \sigma_0,$$

where k_α is the value along the abscissa of the standard normal distribution corresponding to the probability level of $1 - \alpha$. The parameter σ_0 is the standard deviation of the net measurement when no radiation is detected.

The detection limit L_D requires the use of β (not to be confused with the notation for the decision parameter β used throughout the remainder of this text), which is the type II error – the failed detection error. The equation for L_D is

$$L_D = L_C + k_\beta \sigma_D,$$

where k_β, like the k_α value, is to be found on the abscissa of the standard normal distribution and corresponds to the probability level $1 - \beta$, and σ_D is the standard deviation of the net counts at the detection limit L_D. At the detection limit one can ascertain that a signal other than background is present with $100(1 - \beta)\%$ confidence, (the $1 - \beta$ quantity is also called statistical power – the degree to which the null hypothesis is false). Currie also defines an L_Q limit, defined as the determination limit. This quantity is not widely in use and thus its discussion will be omitted.

The limits given by L_C and L_D have long been in use by health physicists. They are typically expressed as

$$L_C = 1.645\sqrt{2}\sqrt{\mu_b} = 2.33\sqrt{\mu_b}$$

and

$$L_D = 2.71 + 4.65\sqrt{\mu_b},$$

respectively (Currie 1968). In both the L_C and L_D equations, the type I and type II errors are both set to 5% (i.e., $\alpha = \beta = 0.05$). In these equations μ_b is the mean background counts. Recall that radiation counts fit a Poisson distribution, and thus $\sigma = \sqrt{\mu}$.

An additional limit called the minimum detectable activity (MDA) is also widely in use. The measure incorporates the efficiency of the detector, and thus acts as a minimum detection limit specific to a detector. The equation for this quantity is

$$MDA = \frac{L_D}{\epsilon \cdot t}$$

such that ϵ is the efficiency of the detector and t is time (Cember, Johnson 2009).

In this research, alternative methods for determining L_C (or a similar expression) will be investigated.

The International Organization for Standards (ISO) 11929 defines the critical level L_C as the decision threshold y^*. Conceptually, this is very similar to the L_C in that it represents an investigatory level. The decision threshold y^* is insufficient to conclude with a level of statistical significance that a signal other than background is present. The equation for the decision threshold is

$$y^* = k_{1-\alpha}\tilde{u}(0),$$

where $k_{1-\alpha}$ is the value along the abscissa of the standard normal distribution corresponding to the probability level of $1 - \alpha$ and $\tilde{u}(0)$ is the standard uncertainty of the background (ISO 11929 2010). The value $k_{1-\alpha}$ is typically set to 1.645, which is the single-tailed value for $1 - \alpha = 0.95$. Likewise, the ISO 11929 (2010) defines the detection limit as the smallest value of the measurand by which one can determine a signal other than background is present with $100(1 - \beta)\%$ confidence. The detection limit, $y^{\#}$, is obtained as the smallest solution to the equation

$$y^{\#} = y^* + k_{1-\beta}\tilde{u}\left(y^{\#}\right).$$

These methods of detection threshold determination are by no means the only available. They are introduced here because they are the most historic (as in the Currie formulas), or they are most often used. For the sake of completeness, Tables 1 and 2 provide alternative ways to determine the decision threshold and the detection limit, respectively.

Table 1 – Various names and sources for the decision threshold

Name	Symbol	Formula (if applicable)	Reference
Decision Threshold	y^*	$y^* = k_{1-\alpha}\tilde{u}(0)$	ISO 11929
Critical Decision Level	L_C	$L_{\mathrm{C}} = 1.645\sqrt{\frac{r_{\mathrm{b}}}{t_{\mathrm{b}}} + \frac{r_{\mathrm{b}}}{t_{\mathrm{g}}}}$	ANSI N42.23
Critical Value	S_C or y_C	$\Pr(\hat{S} > S_C)\,\|X = 0) = \alpha$	MARLAP
Critical Value	$DCGL_w$		MARSSIM Fig. D9
Decision Level	$DL(R_n)$	$DL(R_n) = 1.645\sqrt{R_b\left(\frac{1}{T_b} + \frac{1}{T_g}\right)}$	NUREG 1400
Decision Level	L_C or D_C		ANSI N13.30
Critical Level	L_C	$L_C = 2.33\sqrt{\mu_B}$	Currie

Table 2 – Various names and sources for the detection limit

Name	Symbol	Formula (if applicable)	Reference
Detection Limit	$y^{\#}$	$y^{\#} = y^* + k_{1-\beta}\tilde{u}(y^{\#})$	ISO 11929
Minimum Detectable Value/Concentration	S_D , MDC	$\Pr(\hat{S} \le S\|X = x_D) = \beta$	MARLAP
$0.5\ \mathrm{DCGL_w}$ *	$0.5\ DCGL_w$		MARSSIM Fig. D.9
Minimum Detectable Activity	MDA	$MDA = \frac{\left(2.71 + 3.29\sqrt{R_b T_g\left(1 + \frac{T_g}{T_b}\right)}\right)}{KT_g}$	NUREG 1400
Minimum Detectable Activity	MDA or MDC		ANSI N13.30
Detection Limit	L_D	$L_D = 2.71 + 4.65\sqrt{\mu_B}$	Currie

Prior use of Bayesian statistical tools

The use of Bayesian statistical tools for source detection is far less common, and only recently has it received some attention.

One notable Carnegie Mellon research group tested a Bayesian algorithm for mobile detection (i.e., the detector itself is mobile and is engaged in detecting stationary sources) (Sanquist et al 2008). The background over a large area was characterized using eigenvalue decomposition.

Another research group investigated time-interval data for environmental radiation monitoring (Luo et al 2013). The time-interval analysis uses the time difference between two consecutive pulses (rather that the number of counts in each time interval). This method was found to be advantageous for low radiation levels.

A recent Bayesian model tested a Colorado State University was the Bayesian interaction model (Brogan 2018). This model will be discussed in greater detail later. The model is important because much of the research in this dissertation will be based on the ideas of Brogan.

Frequentist and Bayesian statistics

The use of Bayesian statistics is becoming more popular in many different fields. The whole of Bayesian statistics is rooted in Bayes formula:

$$p(\beta|y) = \frac{p(y|\beta)p(\beta)}{p(y)}.$$

This innocuous looking equation is the foundation to an entirely different school of statistics. It is not the objective of this dissertation to provide a rigorous introduction to Bayesian statistics. However, a short review of the four functions in this equation is provided:

1. The prior function $p(\beta)$: The probability density function (PDF) of the parameter for each $\beta \in \text{Domain}(\beta)$. The prior describes our belief that β represents the true population characteristics expressed by β.
2. The likelihood function (or sampling model) $p(y|\beta)$: This conditional statement describes our belief that y would be the outcome of our experiment given a fixed value for β.
3. The marginal probability function $p(y)$: The probability of observing the data y for all $y \in \text{Domain}(y)$.
4. The posterior function $p(\beta|y)$: This PDF describes our belief that β is the true value, having observed the dataset y, for each value of $\beta \in \text{Domain}(\beta)$.

Since the marginal probability function $p(y)$ is always a constant, Bayes equation is often simplified to the following proportion:

$$p(\beta|y) \propto p(y|\beta)p(\beta).$$

One of the most notable differences between frequentist and Bayesian statistics is interpretation of probability. In frequentist statistics, probability is defined in an a posteriori manner (i.e., it is post-experimental). Probability is the relative frequency that an event will happen over multiple trials or over a long period of time. Probability can be described by the frequency argument, namely that probability equals the relative frequency obtained in a long sequence of repeated measurements, assumed to be performed in an identical manner, physically independent of each other (Gelman et al 2014). In Bayesian statistics, probability is the degree of belief that an event will occur. Hence, it is defined in an a priori manner (i.e., it is pre-experimental). Bayesian statistics does not require an imaginary repeated sampling; it is driven solely by the data.

Both frequentist and Bayesian statistics view the data as randomly generated. The randomly generated data are modeled by a statistical distribution (normal, Poisson, binomial, etc.).

In frequentist statistics, parameters are unknown and fixed. Frequentists assume that there exists a real value for a given parameter in a population. Epistemologically, it is impossible to know this value, (if it were possible, statistics would become unnecessary). Frequentist statistics aims to asymptotically approach the real value through repeated sampling (in addition to calculating the standard deviation based on the amount of sampling). Additionally, a frequentist estimate of the parameter only converges to the true value if there is no bias. The additional assumption of "no bias" is undesirable to Bayesian statisticians; bias can be quantified by addition into the Bayesian equation itself, i.e., $\beta = \hat{\beta} + Bias(\beta)$. The estimator $\hat{\beta}$ is a function of the gathered data such that frequentists assume that $\hat{\beta} = u(y)$. Bayesian estimation makes no such assumption. According to the Bayesian method, the parameter is set equal to a function of the data *as random data,* hence $\beta = u(Y)$. Thus, since the function of a random variable is itself a random variable, β is considered a random variable in Bayesian statistics. As such, the parameter is not fixed but is itself random, and has a probability distribution of its own. The parameters are stated in terms of probability statements. It is for this reason that Bayesian statistics is often seen as more empirical – a primary motivation for Bayesian thinking (Gelman et al 2014).

Another notable difference between frequentist and Bayesian statistics is that frequentist statistics uses hypothesis testing. Entire chapters in introductory statistics texts are devoted to hypothesis testing. Hypothesis testing assumes a fixed parameter and, given that assumed parameter, the probability of observing a given set of data is calculated. If the observation is sufficiently improbable given the assumed value of the parameter, (usually 5%), the hypothesis that the parameter has the previously assumed value is rejected. In Bayesian statistics, the parameter (since it is itself treated as random) is given as a probability distribution as a function of the data and the prior assumptions of the parameter. The probability distribution as a function of the data is a relationship expressed by the Bayes' equation. Hypothesis testing has been

applied to Bayesian inference, but it is usually done in the spirit of synthesis to demonstrate how frequentist and Bayesian statistics are related.

Both frequentist and Bayesian statistics use confidence regions. The regions are a subset of the parameter space that are likely to contain the true value of the parameter. The following definitions are provided by Hoff (2009). A random interval $[l(Y), u(Y)]$ has $100(1-\alpha)\%$ frequentist coverage for β if, before the data are gathered,

$$\Pr(l(Y) < \beta < u(Y)|\beta) = (1-\alpha).$$

An interval $[l(y), u(y)]$, based on the observed data $Y = y$, has $100(1-\alpha)\%$ Bayesian coverage for β if

$$\Pr(l(y) < \beta < u(y)|Y = y) = (1-\alpha).$$

The important difference here is that frequentist coverage is achieved asymptotically, whereas Bayesian coverage intervals are constructed utilizing the observed data.

The prior and the posterior probability functions are functions of the same variable with the important distinction that the posterior is contingent on new information. Because of this feature, the Bayes' formula allows for updating, albeit this is not always true for more complicated Bayesian applications. The posterior becomes the new prior when new information is gathered. The method of updating may be useful for the initial algorithmic training using the training background.

One last difference between frequentist and Bayesian statistics is more esoteric and less known. The class of all Bayes' solutions forms a complete class (Wald 1947). This means that the set of all Bayes' solutions contains the optimal solution. This is one major advantage of Bayesian statistics. The proof furnished by Wald, however, is not a constructive proof – it does not provide an equation or procedure for finding the optimal solution.

These important differences between frequentist and Bayesian statistics are not exhaustive. The motivation for testing Bayesian algorithms for the radiation detection of a weak signal are based on the above arguments, summarized in Table 3.

Table 3 – Summary of Major Differences between Frequentist and Bayesian statistics

Frequentist	Bayesian
Probability: the relative frequency that an event will happen in the long run	Probability: a degree of belief that an event will occur
The data as random	The data as random
Parameters unknown and fixed	Parameters are not fixed but random
$\boldsymbol{f(y|\beta)}$ and test for $\boldsymbol{\beta}$	$p(\beta|y) \propto f(y|\beta)p(\beta)$
Confidence intervals	Credible intervals
Data do not continuously update	Bayesian updating
Is not a “complete class”	Forms a “complete class”

Research scenario vs. field scenario

The method used to detect a potential radiation source against background is context dependent – it matters whether one is in a research scenario or a field scenario. For instance, in the research scenario there are three distinct sets of data (only two are used), but in the field scenario there are two distinct sets of data (both are used). Background means something different in both the research and field scenario. Table 4 is given to provide illumination on these differences.

Table 4 - The research scenario vs. the field scenario

Research Scenario	Field Scenario
Presence of the source is known	Presence of the source is unknown
The purpose of the algorithm is to test the effectiveness of the algorithm.	The purpose of the algorithm is to detect a source.
There are three distinct data sets: (1) the training background, (2) the testing background, and the (3) the testing source.	Two distinct data sets are present: (1) the training background (in the field) and the (2) field counts.
The training background is not necessary for research since the decision threshold is allowed to vary over the range of the other two spectra to form an ROC curve.	The training background (along with the choice of the type I error p) is necessary to determine the decision threshold against which the counts are measured.
Therefore, two spectra are used and compared: the testing background and the testing source.	Therefore, two spectra are also used and compared: the training background and the field counts.
The testing background is compared to the testing source under operationally equivalent conditions, including count time. This is necessary because the testing background is input to the x-axis of the ROC curve.	The training background is compared to the field counts not necessarily under operationally equivalent conditions. They often have different counting times (usually background is longer). This is because the training background is not acting as a control.
The algorithm is run many times to get an average of true and false positives in addition to obtaining these averages over multiple decision thresholds.	The algorithm is run once to determine if there is a source present or if there isn't a source present.
Hence, much longer computation time, but long computation time not an issue.	Hence, shorter computation time, however, is an issue if too long.
The output is an ROC curve.	The output is a binary result of "detect" or "no-detect".

The greatest source of confusion is that in the research scenario the two spectra being compared are *known* background and *known* source counts under operationally equivalent conditions; the background is a control for the experiment. In the field scenario the two spectra

being compared are a *known* background (training background) and an *unknown* status of source presence on the field counts. They are *not* operationally equivalent, and the training background is used to determine the decision threshold. The procedure used in this study will accord itself with the research scenario.

MATERIALS AND METHODS

Data acquisition and analysis

The data used for this research were acquired with a 5 µCi ^{137}Cs source and an open window 2”×2” NaI(Tl) scintillation detector. The source was placed at various distances from the detector (100 cm, 200 cm, and 400 cm) to simulate source strength. The detector resolution was 8.5% at the 662 keV photopeak – with an approximate full width half maximum (FWHM) of 55 keV. The detector was connected to a photomultiplier tube model 2007P (Canberra Industries Inc., Meriden, CT). The preamplification signal was modified by a Lynx Digital Signal Processor (Canberra Industries Inc., Meriden, CT). A spectrum of 1024 channels ranging from 40 – 1700 keV was recorded by the Prospect software (Canberra Industries Inc., Meriden, CT) as the multi-channel analyzer. Data were exported to an excel spreadsheet. Each spectrum contains data for 1 second duration; for any given source-detector geometry, spectra were acquired continuously for over 30 minutes for about 2500 sets of spectra. Previous research has shown that background can change significantly throughout the day at any given location (Fischer 2018). In fact, background has been observed to fluctuate by as much as a factor of three throughout the day, with the highest background occurring at night. All the spectra in this research were taken early afternoon since this is the time of day that the background is the lowest and most stable. These measurements were taken at the Molecular and Radiological Biosciences building at Colorado State University in January of 2018. Figure 1 provides a visual representation of the detector arrangement.

Figure 1 - NaI detector schematic

The Bayesian linear regressions developed in this research each require two sets of measurements: a source and no-source spectra. The full sets of source and no-source spectra cover a duration of about 30 minutes resulting in approximately 2500 individual measurements. Many simulations are needed for the algorithms to obtain results that converge. Usually, 10^4 to 10^5 simulations are needed to obtain converging results. Generating 10^4 to 10^5 iterations from limited data can result in "overtraining" the data. Overtraining means that random effects might take on some significance. In the Bayesian regressions, the data vectors x and y will be of length five, (i.e., $x = (x_1, x_2, x_3, x_4, x_5)$ and $y = (y_1, y_2, y_3, y_4, y_5)$, where each x_i and y_1 is a single measurement). We want to avoid reusing the same vectors. There are a total of 2496 (2500 – 4) consecutive length-five vectors in 2500 measurements. A random sample of 5 is drawn from the total for each iteration to create variation. The sampling assumes that the order in which the measurements are taken is not relevant. The assumption of order irrelevance is an important assumption and will be discussed later. This method of random sampling from the data will be referred to as the random sampling technique.

The R programing language and software is the primary computational tool used for this research. R is an open-source programing language and environment for statistical computing and graphics (The R Foundation 2018). No R packages are used for this research. Certain sections of this paper present coding that is deemed relevant to the discussion. A simplified code is presented in these cases – a coding that is absent some of the necessary but confusing rigors of the code.

Physics review – Spectroscopy and a brief review of gamma interactions

Gamma spectra from a NaI scintillator were used as data. Different aspects of the spectra were utilized for detection based on the physics involved.

Spectroscopy is the aspect of radiation measurement that deals with measuring the energy distribution of the particles being emitted (Tsoulfandis, Landsberger 2015). The energy

distribution is presented as a histogram with the energy of the particle along the abscissa and the number of particles counted at a particular energy along the ordinate and is called a differential energy spectrum (Tsoulfandis, Landsberger 2015). Differential energy spectra can be expressed as a function of $N(E)$ where E is energy and $N(E)$ is the number of particles detected at energy E. In this application, E represents a range of energies; so, more accurately E could be expressed as $E + \Delta E$ for a constant ΔE. The energy interval $E + \Delta E$ is called an energy bin: it is the ith energy bin with minimum energy E_i.

Throughout this research, the full spectrum of gamma energies is utilized for detection. The major advantage of using the entire spectrum of energies to discern the presence of a source is that most sources will increase counts in only a few energy regions. An algorithm can be designed to focus on certain regions depending on the source. The shape of a spectrum is dependent on how the radiation in question interacts with the detector. Since the NaI detector is used to detect gamma rays, an understanding of how gamma rays interact with matter is necessary. Being electrically neutral, gamma rays are difficult to detect directly. Instead, the secondary particles (in this case electrons) are directly detected. The following list offers a very brief review of the three primary ways photons interact with matter and why they are important in this research (Cember, Johnson 2009), (Turner 2007).

(1) Photoelectric effect. The gamma imparts all its energy to an electron, resulting in a monoenergetic electron equal to the energy of the gamma minus the binding energy of the electron (the energy needed to ionize the atom): $h\nu = E_{pe} + \varphi$. In the energy ranges for this research, the binding energy is negligible, thus $h\nu \cong E_{pe}$. The monoenergetic peak is called the *photopeak* in a gamma spectrum. For a ^{137}Cs source, the photopeak is at 662 keV.

(2) Compton Scattering. In this interaction the incoming gamma ray imparts only some of its energy to an electron, resulting in a scattered gamma ray of less energy and an electron. The scattered gamma and electron occupy a continuum of energies. The electron energy

ranges from 0 to a maximum energy less than the photopeak energy. The maximum energy of the electron is called the *Compton edge.* The Compton edge in the ^{137}Cs spectrum is 478 keV.

(3) Pair Production: In this interaction a high energy gamma produces an electron/positron pair. The minimum energy required for this reaction is 1.022 MeV. This interaction is not relevant in this research because the energy is higher than the energy released in the decay of ^{137}Cs.

In this research the photopeak and the Compton edge are used extensively in the Bayesian algorithms to be introduced.

Procedure

The development and testing of a Bayesian model are often not as simple as a frequentist model. Frequentist methods fit a probability distribution to background radiation and perform hypothesis testing. The distributions most commonly employed in frequentist methods are the normal distribution and (less frequently) the Poisson distribution. Hypothesis testing is performed on a simple parameter (usually the mean). In a Bayesian algorithm, the parameter is not always an obvious function of the data; it may not even resemble a known physical measurement. Additionally, Bayesian computation may require Monte Carlo approximations. The method used by this research (or the application of the research) has been separated into seven steps. These steps are not necessarily performed in order, and some of these steps can even be omitted.

(1) <u>Determine a statistical model to describe radiation phenomena.</u> Due to its stochastic nature, random variables are used to represent radiation phenomena. Associated with any random variable is a probability distribution, thus any random variable will have associated with it a probability density function (PDF) and/or a cumulative distribution function (CDF), (recall that the PDF is the derivative of the CDF). The three most common statistical distributions used to describe radiation phenomena (at least for frequentist statistics) are the binomial distribution,

the Poisson distribution, and the normal distribution. Complexity can be added when certain parameters of these distributions are made to be more complicated functions. For example, the parameter λ in the Poisson distribution, representing the mean and variance, could be a linear regression of multiple variables, (i.e., $\lambda = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$); (of course, one must examine whether the application of additional complexity is warranted). The chosen model can be expressed by the variable *M*.

(2) Gather information about background. This process is known as "training the algorithm." The background is called the training background. After choosing a statistical model for the background radiation, the background data is used to "train" the parameters of the model towards certain values. This essentially means the same thing as convergence. The rate of convergence depends on the amount of data and the model being used. This process is also known a machine learning – a process very similar and often identical to Bayesian analysis. Whereas step one provides the "form" or the "structure" of the model, step two provides the "matter" by furnishing data to fill the model. Taken together, steps one and two construct a Bayesian model for background radiation. Symbolically, this can be expressed as (*training background* | *M*). The primary purpose of the training background is to determine the decision threshold. This step is necessary in the field scenario; however, in the research scenario, this step is omitted because the decision threshold is not given a set value. Instead, it is allowed to vary over the domain of the decision parameter.

(3) Choose a probability to determine desired success/fail rate. Traditionally this value is denoted as α; in this section, it shall be referred to as p such that $p \in [0,1]$. Statistically, this value refers to the type I error probability; in radiation protection it is called the false positive rate. This value has been typically set to 0.05 and has been aptly described as a "happy intuition" by the pioneering theorists (Neyman, Pearson 1933); of course, this value is quite subjective. Since that time, the value 0.05 has remained widely in use. However, in some applications a false positive probability of 0.05 may be unacceptably large. The subjectivity in

setting a value of 0.05 is somewhat inevitable. The subjectivity is entirely circumvented, however, in our use of Receiver Operator Characteristic (ROC) curves as a performance metric. When graphing ROC curves, the p value ranges over its entire domain [0,1] (just as β ranges over its domain). For this reason, this step is also omitted in the research scenario. ROC curves will be discussed in a subsequent section.

(4) Choose a decision threshold. The decision threshold is an assigned value based on the training background that determines above (or below) which value a source is determined to be present. This can be expressed as $\beta^* = f(p, training\ background\ |M)$ where f is a function. Traditionally, this value is $y^*$and is defined as $\alpha = p(y > y^*|background, M)$ where *background* are the background counts, and the model *M* is the normal distribution. In the field scenario, β^* is functionally dependent on the chosen false probability rate and the training background. In the research scenario, β^* ranges over multiple values, (that is why steps 2 and 3 are not needed in the research scenario).

(5) Choose a decision parameter. The decision parameter is a function of the new data; it can be expressed as $\tilde{\beta} = g(new\ data\ |M)$ where *g* is a function. It is essential that β^* and $\tilde{\beta}$ have the same domain so that both variables are in the same set, (i.e., $\mathrm{Domain}(\beta^*) = \mathrm{Domain}(\tilde{\beta})$). In this case, the domain will be the positive real numbers. The decision parameter is chosen from a set of different parameters, $\tilde{\beta} \in \{\beta_1, \dots, \beta_n\}$ for n different possible parameters. In the linear regression models presented in this research, there are three different parameters tested as a decision parameters per model. Traditionally the decision parameter is simply the number of counts y when testing for the presence of a source.

(6) Apply decision criterion to determine if a detection event has occurred. The decision criterion is expresses as $\tilde{\beta} > \beta^*$, (or perhaps $\tilde{\beta} < \beta^*$). This is the criterion by which a decision is made such that only two possible scenarios present themselves: either there is a positive indication of $\tilde{\beta} > \beta^*$ and a source is said to be present, or there is a negative indication of $\tilde{\beta} \leq$

β^* and a source is said to not be present. In the traditional model, this criterion is given by $y > y^*$ (ISO 11929 2010).

It is important to understand that a detection event is understood to mean, "the observed data would be observed $100 \times p$ % of the time given the current measurement of training background." The value $100 \times p$ % is small enough such that we suspect the present of something other than background. This is typically referred to as an investigatory level of detection. With this additional nuance in mind, and for the sake of simplicity, we will continue to say that a "detection event has occurred," or a "source is present," whenever a positive indication occurs.

(7) Test this method of detection against other known methods. The new method will always be tested against the traditional method of $y > y^*$ as this is the method most often employed in the field. The ISO 11929 (2010) method shall henceforth be referred to as the "traditional method."

Assumptions

The following assumptions are made in our experiments.

(1) The random variables are independent. Since the Bayesian modeling is of radiation phenomena, it is assumed that any emission of radiation is independent of any previous or subsequent emission of radiation. This independence allows for the following mathematical simplification:

$$p(y_1, y_2, \dots, y_n|\beta) = p_{Y_1}(y_1|\beta)p_{Y_2}(y_2|\beta) \dots p_{Y_n}(y_n|\beta) = \prod_{i=1}^{n} p_{Y_i}(y_i|\beta),$$

such that each y_i is a measurement of radiation counts for n measurements given some parameter β. In other words, any given Y_i gives no information about Y_j for all $i \neq j$. Additionally, since each Y_i is generated in a similar fashion from a common process, it can further be assumed that the marginal densities P_{Y_i} are all identical. This gives a further simplification of

$$p_{Y_1}(y_1|\beta)p_{Y_2}(y_2|\beta)\dots p_{Y_n}(y_n|\beta) = p(y_1|\beta)p(y_2|\beta)\dots p(y_n|\beta) = \prod_{i=1}^{n} p(y_i|\beta).$$

The mathematical shorthand for this is

$$Y_1, \dots, Y_n|\beta \sim \text{i.i.d.}\, p(y|\beta),$$

(Hoff 2009).

(2) The data are exchangeable. The data are exchangeable if any permutation of the data yield the same result. Thus, if

$$p(y_1, \dots, y_n) = p\left(y_{\pi_1}, \dots, y_{\pi_n}\right)$$

for all permutations π of $\{1, \dots, n\}$, then $Y_1, \dots, Y_n$ are exchangeable (Hoff 2009). Essentially, the data are exchangeable if the subscript labels of $Y_1, \dots, Y_n$ convey no meaning; they are merely an arbitrary indexing of an unordered set.

The first assumption holds for all radiation phenomena in any scenario; it does not matter what is the type of radiation, its energy distribution, or its activity or changing activity. Every nuclear decay event is completely independent. The second assumption, however, does not hold in general for radiation phenomenon, but only in those scenarios where total activity does not change $\left(\frac{dA}{dt} = 0\right)$ for all energies. These two statements are in fact equivalent: the data are exchangeable if and only if there is not a change in counts over time. This is of course an inaccurate generalization. Background is known to fluctuate over the day (Fischer 2018). There are times during the day where background counts are relatively stable – usually around midday. The data used for this research were taken during this time in a 30-minute window. Since this research is not concerned with the study of changing background, this assumption is both an acceptable and desirable simplification. Additionally, this assumption is also necessary if we are to justify the random sampling technique discussed earlier.

Traditional method of radiation detection

In this research, all alternative detection algorithms are compared against the traditional method of detection. The method used as the traditional method is based on the ISO 11929 terminology; the decision threshold denoted as y^* will be used. The derivation of y^* is based on the equation:

$$P(y > y^* | H_0 \text{ is true}) = \alpha.$$

The null hypothesis H_0 is that hypothesis that assumes that no source is present. For this method we assume a normal distribution. The probability α is the error of the first kind – the false positive probability given no source is present. A graphical representation of this concept is given in Figure 2 (Meengs 2018).

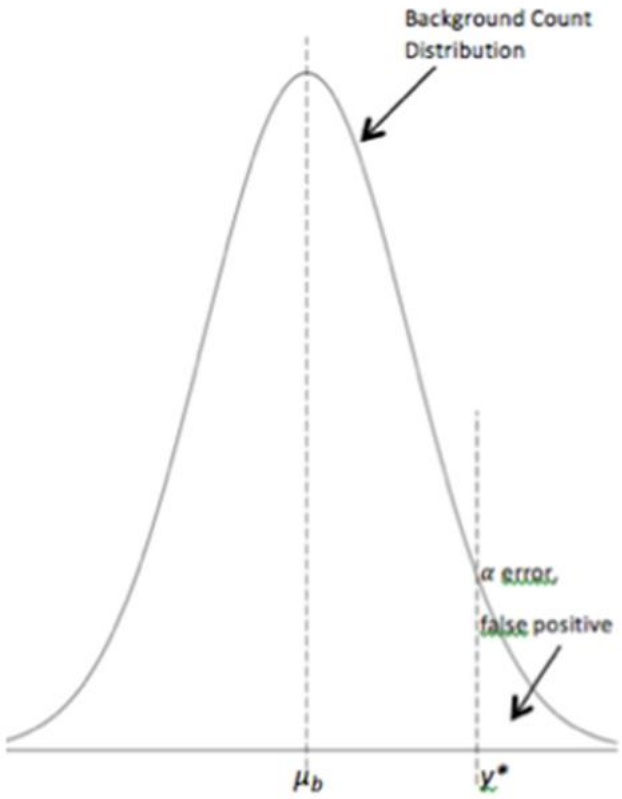

Figure 2 - Background distribution with decision threshold and type I error

The calculation of y^* under these assumptions requires the use of numerical methods. The PDF for the normal distribution is

$$P(Y = y) = \frac{1}{\sigma_b\sqrt{2\pi}} \exp\left[-\left(\frac{1}{2\sigma_b^2}\right)(y - \mu_b)^2\right].$$

The integral to this function has no closed form and will thus result in an error function if an attempt is made to solve this using any software. A standard normal distribution table (z table) must be consulted to solve for y^*. In the ISO 11929 (2010) standard, the solution is given by $y^* = k_\alpha \sigma_b$. The R code will do this with the $\mathrm{qnorm}(\alpha, \mu, \sigma)$ function.

As with the β parameters being researched, the calculation of y^* given an α value and background is only necessary in the field scenario. This research does not require particular calculations of y^* as a derivation of a desired α value. Instead, y^* is allowed to vary over a given range as each point of the ROC curve is calculated. Specifically, y^* will vary over the given parameter space, (in this case just the counts y), such that $y^* \in [\min(y), \max(y)]$. The decision threshold y^* will vary discretely over this range in which the false and true positives will be calculated for every possible value for y^*. This procedure will yield an ROC curve. The mathematical process by which an ROC curve is generated will be discussed in the next section.

ROC curves as the metric for determining optimal detection

For both the spectral optimization algorithms and the Bayesian algorithms, ROC curves will be used to determine the success of the optimization. An ROC curve is a parametric curve (i.e., the independent variable is not expressed on either the abscissa or the ordinate). Along the abscissa of the ROC curve are the false positive rates for a given value of the decision threshold β^*; along the ordinate of the curve are the true positive rates for the same decision threshold. The rates are not frequentist probabilities, but actual calculated rates of occurrence as the algorithm is executed many times. Expressed mathematically, an ROC curve is the set of

all points $\left(\text{false positve}(\beta^*), \text{true positive}(\beta^*)\right)$ such that β^* ranges over its entire parameter space, expressed as $\beta^* \in [\min(\beta), \max(\beta)]$.

In Figure 3, two distributions are shown.

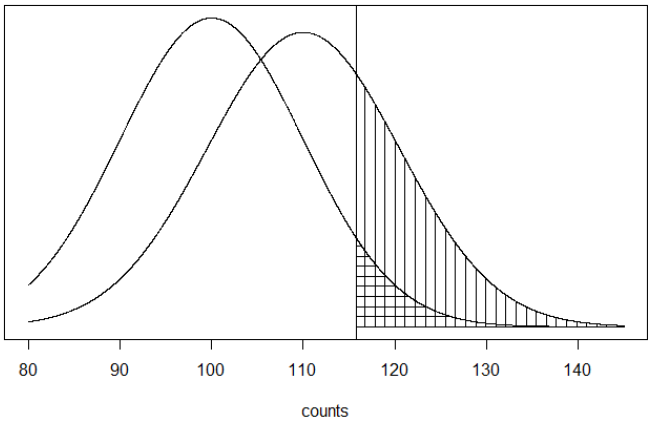

Figure 3 - Graphical representation of a point along the ROC curve given a value for β^*

In this example, the curve to the left with $\mu = 100$ is background, and the curve to the right with $\mu = 110$ is the background plus some source (giving a net count of 10). A particular value for the decision threshold is given by the vertical line. The areas under the curves to the right of the decision threshold represents the false positive rate and the true positive rate for the background (left) and source (right) curves, respectively. This is because the areas to the right of the vertical line represent events where the decision parameter is greater than the decision threshold (i.e., $\tilde{\beta} > \beta^*$). Specifically, the vertically shaded region is the true positive rate, and the horizontally shaded region is the false positive rate for that given decision threshold. A particular point along the ROC curve is then given by $(x, y) = (\text{Area}(horizontal\ region)\ , \text{Area}(vertical\ region))$. As the decision threshold β^* (given by the vertical line) sweeps across both distributions from left to right, an ROC curve is traced from the

points (1,1) to (0,0). The example provided uses counts to illustrate the formation of an ROC curve, but the concept remains unchanged for comparing any two parameters.

Four examples of ROC curves are given in Figure 4.

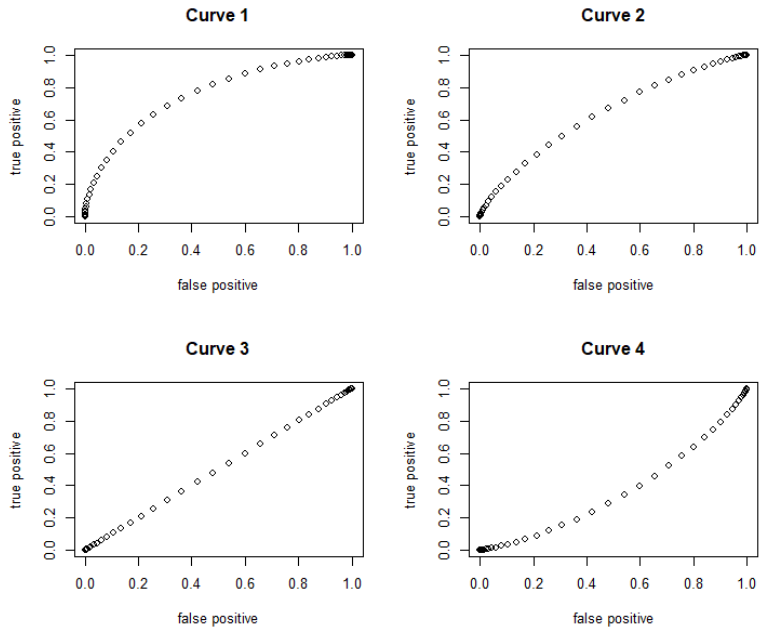

Figure 4 - Four examples of ROC curves

The first curve in the figure shows the highest true positive rates for any given false positive rate. In general, the concave up shape is desirable. An increased concave upward shape corresponds to better detection. Thus, curve 1 expresses a better job at detecting a source than curve 2. The 3rd curve is simply the $y = x$ line. This straight diagonal line corresponds to a complete lack of detection ability as the false positives are equal to the true positives; the detection of a source is essentially random. The fourth curve is concave downward. As shown, this would mean that the false positives outnumber the true positives. One may naively think that this is worse than random. This, however, is not so, as this nevertheless demonstrates an ability to discriminate between the two spectra. What this type of curve indicates is that the decision criterion was chosen incorrectly. If the decision criterion was $\beta > \beta^*$, then an ROC

curve like curve four would indicate that the correct criterion was $\beta < \beta^*$. After reversing the criterion, the curve would invert itself in the concave up configuration. In cases where this occurs, however, the algorithm is not repeated and instead it is noted that the reverse criterion is to be used. The curve is then inverted for the sake of consistency.

SPECTRAL OPTIMIZATION ALGORITHM

The research was divided into two distinct parts. In the first part, the spectral optimization algorithm was examined. The purpose of the spectral optimization algorithm is to systematically discriminate between parts of a spectrum which likely contribute to a source signal from those signals which likely contribute to noise. If successful, spectral streamlining (i.e., the stripping away of all noise from the spectrum) increases the likelihood of a successful detection. In the next section, different Bayesian linear regressions were tested against the traditional method of detection.

Background

Much of the previous research, including that of the Bayesian interaction model (Brogan 2018), utilized only the sum of counts from all energies. This constitutes a great simplification of the data; consequently, information is lost. This research endeavors to utilize spectral information.

Some simplification, however, is useful. The data being used for this research originated from a NaI detector – a detector with an 8.5% resolution about the 662 keV photopeak. The data were originally partitioned into 1024 energy bins. This division is excessively fine given the resolution of the detector. It therefore suffices to have an energy partitioning commensurate with the full width half maximum (FWHM) of the detector, about 56 keV (0.085×662 keV). The FWHM is the highest resolution achievable before information is lost due to the measurement uncertainties of the detector. Therefore, with the NaI detector, an energy partitioning into 50 keV width bins is recommended.

Figure 5 gives an example of two energy spectra – a background spectrum and a source spectrum (the source spectrum including both source and background).

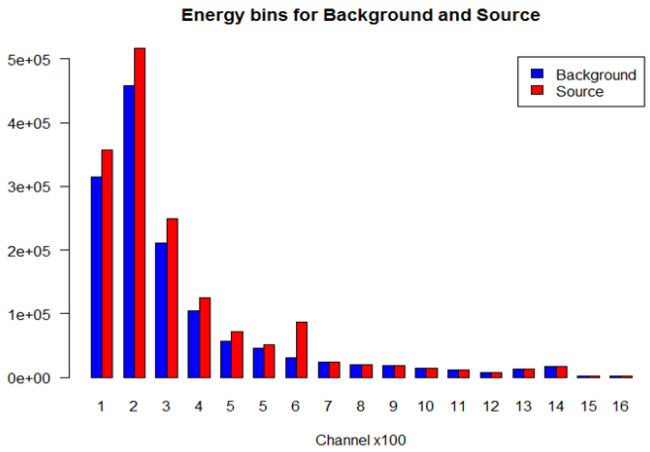

Figure 5 - Comparison of energy bins for background and source (^{137}Cs)

The spectra are from background and a weak ^{137}Cs source, respectively. In this example, the bin widths are set to 100 keV where the bin number is the minimum energy in hundreds of keV (the 0th bin of <100 keV is not included). Figure 5 demonstrates that bins 7-16 (700 keV – 1600 keV) are pure noise, and negatively contribute to the signal's detection. Thus, the removal of bins 7-16 will increase detection sensitivity. Bin 6 contributes considerably to detection sensitivity and thus must be retained. The relative source detection contribution of bins 1-5, however, are less obvious. Bins 1-5 contribute to both signal detection and noise. The question is: do these bins contribute more to source detection, or do they contribute more to noise? Or, more broadly: is there a mathematical quantity that can be used to predict whether a bin positively or negatively contributes to source detection? If not, is there an algorithm that can be employed to ensure that all the bins positively contributing to a detection are lumped together, and all bins contributing to noise are excluded?

In the two subsequent sections, the theory behind a spectral optimization algorithm is discussed, followed by a discussion of how such an algorithm is developed.

Theory I – Obtaining a metric for bin optimization

Previous research has shown that the quantity $\frac{\Delta\mu}{\sigma}$ is an accurate estimator of relative source detection sensitivity (Meengs 2018). This quantity is also known as the Cohen's d measure of statistical power (Cohen 1988). In this application, $\frac{\Delta\mu}{\sigma}$ is defined as

$$\frac{\text{average gross counts} - \text{average background counts}}{\text{standard deviation of background counts}}.$$

As mentioned earlier, $\frac{\Delta\mu}{\sigma}$ is a relative measure of an energy bin's detection sensitivity. The actual criterion for deciding whether one bin e_1 is better at detecting a source than another bin e_2 is given by the following relation:

$$\left(\frac{\Delta\mu_1}{\sigma_1}\right)_{e_1} > \left(\frac{\Delta\mu_2}{\sigma_2}\right)_{e_2}.$$

Since radiation phenomena obey Poisson statistics, the following simplification is made: $\sigma = \sqrt{b}$, where b is the mean background. Following substitution, the equation becomes

$$\left(\frac{\Delta\mu_1}{\sqrt{b}_1}\right)_{e_1} > \left(\frac{\Delta\mu_2}{\sqrt{b}_2}\right)_{e_2}.$$

Previous research has explored the idea of adding or subtracting energy bins (Meengs 2018) and determining how these operations effect detection. If an energy bin e' is combined with the energy bin e the resulting $\frac{\Delta\mu}{\sigma}$ for the energy bin $e \cup e'$ is expressed as

$$\frac{\Delta\mu+\Delta\mu'}{\sqrt{b+b'}},$$

where $\Delta\mu$ is the average net counts $(g - b)$ such that g and b are the mean gross counts and mean background counts, respectively. For this combination to be advantageous, it must be that

$$\frac{\Delta\mu+\Delta\mu'}{\sqrt{b+b'}} > \frac{\Delta\mu}{\sqrt{b}}.$$

Following rearrangement, the following relationship is obtained:

$$\frac{\Delta\mu+\Delta\mu'}{\Delta\mu} > \sqrt{\frac{b+b'}{b}}.$$

This equation is an iterative method for combining bins in the most efficient manner. Essentially, if the relative increase of the net counts is greater than the square root of the relative increase in the background, then the combination is advantageous. In other words, when the above criterion is met for adding a second bin (expressed by the primed values) to an initial bin (expressed by the unprimed values), then the combined bins (the union of the primed and unprimed bin) will result in a higher $\frac{\Delta\mu}{\sigma}$ value, thus better detection sensitivity. Henceforth, this relation will be called the "additive criterion" for combining energy bins. An important assumption with this criterion is that the energy bins are disjoint (i.e., $e \cap e' = \emptyset$).

There is also a subtractive iterative method given by:

$$\frac{\Delta\mu - \Delta\mu'}{\Delta\mu} > \sqrt{\frac{b - b'}{b}}$$

(Meengs 2018). In this case, if the relative decrease in net counts is greater than the square root of the relative decrease of the background, then subtracting the energy bin e' from energy bin e is advantageous. This criterion assumes that the subtracted bin is a subset of the initial bin, (i.e., $e' \subset e$). In this research, however, only the additive criterion is used.

Theory II – Towards finding the optimal combination of energy bins in a spectrum

As previously discussed, a spectrum of 1024 channels is unnecessary given the resolution of the NaI detector. Several channels can be combined to create energy bins. These bins should have a width commensurate with the FWHM of the detector, (e.g., a bin width of 50 keV). The set of all energies (i.e., the entire spectrum) can be expressed as the set S. The entire spectrum S is partitioned into n energy bins to form a collection of sets such that $S = \{e_1, \ldots, e_n\}$, where each e_i is a bin of equal energy width. This simplification is not necessary for the proper functioning of the algorithm because the quantity $\frac{\Delta\mu}{\sigma}$ is not functionally dependent on bin width (i.e., the range of energies contained in a single energy bin). Thus, equal bin width is not a

necessary condition, but it provides a convenient simplification. Moreover, any non-arbitrary partitioning of bins into unequal width might presuppose some "inadmissible" knowledge about the source being detected, (i.e., having a larger bin about the photopeak or the Compton edge etc.). One exception would be the use of logarithmic scaling instead of linear scaling (i.e., logarithmically equal bin size vs linearly equal bin size) (Sanquist et al 2008). These possibilities, however, are not explored in this research. Additionally, each energy bin e_i will have two counts associated with it: the gross counts g_i (containing a source) and the background b_i counts (containing no source).

When the spectrum S is partitioned into n energy bins, the following properties hold:

(1) $e_i \cap e_j = \emptyset$ for all $i \neq j$ (i.e., they are disjoint) and

(2) $\bigcup_{i=1}^{n} e_i = S$.

In other words, S is a partitioned set. The power set of S, given by $\mathcal{P}(S)$, is the set of all possible combinations of the e_i subsets in S. The power set of S is much larger, with a cardinality of 2^n for n bins. This means that there are 2^n different ways to combine the bins e_i in the set S.

The hypothesis is that there exists a member of the power set of S, $\mathcal{O} \subset \mathcal{P}(S)$, that is a union of all the bins that positively contribute to source detection. This is essentially an optimization problem. The idea is that if any energy bin is subtracted from $\mathcal{O}$, then the signal will weaken; additionally, if any bin is added to $\mathcal{O}$, the signal will also weaken. Thus, the set $\mathcal{O}$ will have the largest $\frac{\Delta\mu}{\sigma}$ in the power set $\mathcal{P}(S)$. The set $\mathcal{O}$ will naturally have the following properties:

(1) $\mathcal{O} \subset \mathcal{P}(S)$

(2) $\mathcal{O} \cap \bar{\mathcal{O}} = \emptyset$ and

(3) $\mathcal{O} \cup \bar{\mathcal{O}} = \mathcal{P}(S)$.

Development I – Mathematical construction

This section is an attempt to construct the set $\mathcal{O}$.

Step 1: Calculate $\frac{\Delta\mu}{\sigma}$ for each energy bin e. Recall that the spectrum S can be expressed as the disjoint partitioning of energy bins e_i for n bins as

$$S = \{e_i | 1 \leq i \leq n\}.$$

Define the $\frac{\Delta\mu}{\sigma}$ function as

$$D(e_i) = \frac{g_i - b_i}{\sqrt{b_i}},$$

where the function D maps a set to a non-negative real number. A union of bins is also defined under D as

$$D(\bigcup e_i) = \frac{(\sum g_i - \sum b_i)}{\sqrt{\sum b_i}}.$$

The following ordered set expressing the $\frac{\Delta\mu}{\sigma}$ of each bin e_i can now be calculated:

$$S_D = \{D(e_1), \ldots, D(e_n)\}.$$

Note that $D(S) = \frac{\Delta\mu}{\sigma}$ for the gross counts (which is a number) and thus $D(S) \neq S_D$ as defined, (which is a set).

Step 2: Permute the ordered set S_D in descending order. In R, the command to sort a vector in increasing (or decreasing) order is the sort() function. The permutation will be expressed as π.

$$\text{sort}(S_D, \text{decreasing} = \text{TRUE}) = (S_D)_\pi = \{D(e)_j | 1 \leq j \leq n\}$$

where j is the new ordering under π such that for all $j, D(e)_j > D(e)_{j+1}$. The energy bins must also undergo the same permutation:

$$\mathrm{S}_\pi = \{e_j | 1 \leq j \leq n\}.$$

The idea is that there exists a clean partition of S_π such that $\mathcal{O} = \bigcup_{j=1}^{k} e_j$ for some k.

Step 3: The set $\mathcal{O}$ must not be the null set (i.e., it must contain at least one bin e) so that detection is possible. Naturally, $\mathcal{O}$ must minimally contain $e_{j=1}$ as the first (and perhaps only) member of its set.

Step 4: For the equation $\mathcal{O} = \bigcup_{i=1}^{k} e_{\pi_i}$, use the additive criterion to see if the addition of $e_{\pi_{k+1}}$ increases the $\frac{\Delta\mu}{\sigma}$ of $\mathcal{O}$. If

$$D\left(\bigcup_{j=1}^{k+1} e_j\right) > D\left(\bigcup_{j=1}^{k} e_j\right),$$

then the bin e_{j+1} is also included in the set $\mathcal{O}$. This pattern continues until the above criterion is no longer met. Therefore, $\mathcal{O}$ is defined such that

$$\mathcal{O} = \bigcup_{j=1}^{k} e_j,$$

where $k + 1$ is the lowest positive integer that does not satisfy the additive criterion.

To aid in explaining why this works, the additive criterion can be written in equation form:

$$y = \frac{\Delta\mu + \Delta\mu'}{\Delta\mu} - \sqrt{\frac{b+b'}{b}}.$$

The variable y is an expression of change in detection ability. If $y > 0$ then the criterion is met, and the detection ability is increased. Likewise, if $y < 0$ the detection ability is decreased. This can be expressed in terms of $\frac{\Delta\mu}{\sigma}$. The following substitution can be used for $\frac{\Delta\mu}{\sigma}$:

$$x = \frac{\Delta\mu'}{\sqrt{b'}}.$$

After substituting $x\sqrt{b'}$ for $\Delta\mu'$, the following linear equation is obtained:

$$y = \left(\frac{\sqrt{b'}}{\Delta\mu}\right) x + \left(1 - \sqrt{\frac{b+b'}{b}}\right).$$

The slope term $\frac{\sqrt{b'}}{\Delta\mu}$ is always positive and thus the equation is continuously increasing. For any given b', there is always a positive correlation between $x\left(\frac{\Delta\mu}{\sigma}\right)$ and y. For any x value that gives $y < 0$, any $x' < x$ will also give $y < 0$ when the background b' is held constant. The same result

is obtained when $\Delta\mu'$ is held constant. When the substitution $b' = \frac{\Delta\mu'^2}{x^2}$ is made (again from $x = \frac{\Delta\mu'}{\sqrt{b'}}$), the following equation is obtained:

$$y = 1 + \frac{\Delta\mu'}{\Delta\mu} - \sqrt{1 + \left(\frac{\Delta\mu'^2}{b}\right)\frac{1}{x^2}}.$$

Though less obvious, this function is also continuously increasing for any given value of $\Delta\mu'$. A coding example for the optimization algorithm is provided in Appendix A.

Evaluation and additional ideas

After running the algorithm several times using a known background and source, it has been verified that the algorithm works – that it successfully identifies which bins positively contribute to a signal when a source is present. Yet, the algorithm attempts to do the same when only background noise is present. This attempt to optimize mere noise when a source is not present has the unfortunate effect of rendering the algorithm a poor method for detection. Figure 6 provides the results for the optimization algorithm. The sub-par performance of the algorithm compared to the traditional method is quite evident.

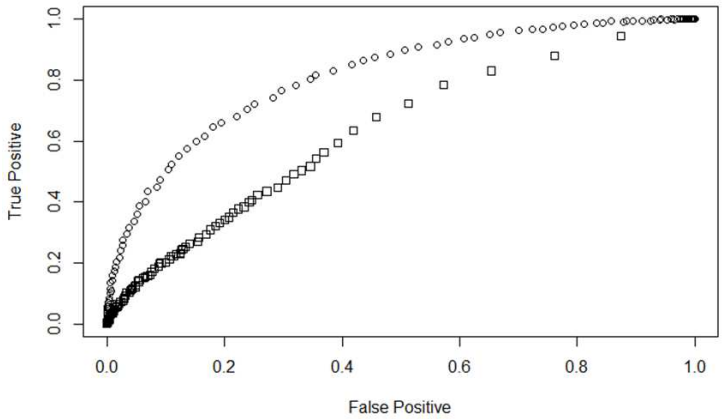

Figure 6 - ROC curves comparing the traditional method with the optimization algorithm

Although the algorithm did not provide an improvement in detection, the development of the algorithm yielded some insights. Redemption may be found in the fact that optimized noise looks very different than an optimized source. These differences are:

(1) *When a source is present*, the bins contributing to a signal are few, and the algorithm naturally chooses these bins. Typically, the bins near the photopeak region are chosen and all other bins are discarded. This results in a *small, optimized bin*. Additionally, the same bin(s) are *repeated* every time the algorithm is executed.

(2) *When a source is not present*, the optimization algorithm attempts to optimize noise. Since the statistical fluctuation is random, the algorithm will choose about half of the bins every time. This results in a *large, optimized bin* when there is not a source present. Additionally, every time the algorithm is executed the "optimized" bins will be *random*. Therefore, when only noise is present, a random selection of about half of the bins will be chosen every time.

It would not be difficult to program a computer to recognize these differences; (indeed, this is precisely what a human observer does when they look at a spectrum). Further development, however, is beyond the scope of this research. Instead, research ideas to make further progress are presented.

A future researcher might want to further characterize the statistical nature of an optimized background. The following research questions are presented.

Question 1: If there is a total of n bins, is the average indeed $\frac{n}{2}$, as expected?

Question 2: How is the number of optimized bins distributed? Is it a binomial distribution (or something similar) with average $\frac{n}{2}$?

Question 3: What is the standard deviation of this distribution?

Once these questions are answered, it may be possible to define a sufficiently small positive integer b_{max} (where $b_{max} < \frac{n}{2}$) such that the probability that b_{max} or less bins are chosen by the optimization algorithm if only noise is present is less than some false positive rate α:

$$p(b \leq b_{max}) = \alpha,$$

where b is the number of optimized bins. The event $b \leq b_{max}$ could be considered a detection event. Further, the detection could be specific to a particular isotope depending on which bin(s) are optimized when a detection is triggered. Additional theories and methodologies could be developed in the future that could lead to other methods of optimizing detection.

PRESENTATION OF THE BAYESIAN MODELS

Background

One of the goals of this research is to offer an extension to the Bayesian interaction model using information from the entire spectrum. The Bayesian interaction model (Brogan 2018) is reproduced below:

$$Count_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = a + \gamma_i \times SD5_i + b_{BkSc_i} \times Bk_{Sc_i}$$

$$\gamma_i = b_{SD5} + b_{BkSc,SD5} \times Bk_{Sc_i}.$$

The priors used for the algorithm were:

$$\alpha \sim Normal(600,10)$$

$$b_{SD5} \sim Normal(0,1)$$

$$b_{BkSc_i} \sim Normal(0,1)$$

$$b_{BkSc,SD5} \sim Normal(0,1)$$

$$\sigma \sim Uniform(0,50).$$

The interaction feature of this model is not readily obvious given how the equation is presented. When the γ_i parameter is substituted into the μ_i equation the following is obtained:

$$\mu_i = a + b_{SD5}(SD5) + b_{BkSc_i}(Bk_{Sc_i}) + b_{BkSc,SD5}(Bk_{Sc_i})(SD5).$$

Here the interaction term (the last term) is more obvious. The equation now follows the general linear regression form:

$$\mu_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2.$$

The last term is a mixed or quadratic term. This added level of complexity is warranted if there is evidence that the x_1 and x_2 terms are not independent. In this model, the terms were set as

$$x_1 = SD5 \text{ and}$$

$$x_2 = Bk_{Sc_i} = \mathbb{I}(background).$$

The value SD5 is the sample standard deviation calculated over 5 measurements and is used as a predictor of the total counts. The function $\mathbb{I}(background)$ is an indicator function that is set equal to one if the measurement is background (i.e., more specifically, if the measurement taken is the "background" used to train the algorithm – the training background). The value is set to zero otherwise. The statistic SD5 is used as a predictor for the average counts μ. Because one of the predictors is an indicator function ($\mathbb{I}(background) \in \{0,1\}$), it has a parameter space of size two. Therefore, the variables could be interpreted as follows:

$\beta_0 + \beta_2 =$ intercept for training data

$\beta_0 =$ intercept for test data

$\beta_1 + \beta_{12} =$ slope for training data

$\beta_1 =$ slope for test data

It therefore follows that the parameter β_2 represents the difference in intercept between test data and training data, and the parameter β_{12} represents the difference between the slope of SD5 for the test data and the training data.

The model could be recast as two linear regressions:

$$Y_b \sim Normal(\beta_{b0} + \beta_{b1} x_b\ , \sigma)$$

$$Y_s \sim Normal(\beta_{s0} + \beta_{s1} x_s, \sigma),$$

such that

$\beta_{bo} =$ intercept for training data

$\beta_{b1} =$ slope for training data

$\beta_{s0} =$ intercept for test data

$\beta_{s1} =$ slope for test data.

The decision parameter used in the Bayesian interaction model was the regression coefficients for SD5. Specifically, they are $\gamma_b = \beta_1 + \beta_{12}$ for the training data and $\gamma_s = \beta_1$ for the test data (Brogan 2018). Therefore, the decision criterion was $\beta_1 > \beta_1 + \beta_{12}$. The probability of

the criterion is calculated and compared to the value Pr*, where Pr* is the decision threshold. A source is determined to be present if the criterion $\beta_1 > \beta_1 + \beta_{12}$ is satisfied to within some probability. Equivalently, this can be expressed as $\beta_1 > \beta_0$ for the newly recast version of the model.

These models are mathematically equivalent. The purpose of demonstrating this equivalency is to show that the models tested in this research are a logical extension of the interaction model. In this research, two linear regression models are used for both the source and no-source counts.

The value chosen for the predictor, SD5, is not the only logical choice. Theoretically, x_1 can be set to any statistical value. Still, more regressors could be added if an additional level of complexity seems warranted. Any one of these regressors could be used as decision parameters.

In both models, σ is a nuisance parameter. It is called a nuisance parameter because there is no interest in making inferences with this parameter, however, it is required to construct a normal model. Additionally, as previously discussed, $\sigma = \sqrt{\mu}$. Hence, these two parameters in the normal model, μ and σ, are not independent. It is therefore desirable to regard the σ parameter as a nuisance parameter and subsequently ignore it. One way to do this would be to reduce the normal model to $Y \sim N(\mu\,, \sigma = \sqrt{\mu})$. Depending on the algorithm being tested, μ will be set equal to a different linear regression.

Introduction

This research will consider several different linear regression models. Even though the simplest model for radiation is the Poisson model (or the binomial model, due to having only one parameter), we will consider only the normal model. This is because the normal linear regression is well known. The normal model is used as an approximation of the Poisson model

at high enough counts. This relationship will simplify what is normally a two-parameter model into a one parameter model since the variance is determined and is in fact equal to the mean.

The general form of the linear regression model will be the simplest of all linear regression models: consisting of only an intercept and slope term. The general form of the model is given here:

$$Y \sim N(\beta_0 + \beta_1 x\,, \sigma)$$

$$\beta_0 \sim N(\mu_0, \tau_0)$$

$$\beta_1 \sim N(\mu_1, \tau_1)$$

$\sigma^2 = \beta_0 + \beta_1 x$ (for gross counts)

$\sigma^2 = (\beta_0 + \beta_1 x) + \mu_{tr}$ (for net counts).

We will set $\sigma^2 = \beta_0 + \beta_1 x$ where gross counts are used and $\sigma^2 = (\beta_0 + \beta_1 x) + \mu_{tr}$ where net counts are used; μ_{tr} is the mean of the training background. The data terms y and x are vectors of length five and, each are taken from different energy bins of the spectra. The reason they are length five is to simulate five seconds of data collection. There is nothing inherently special about 5 seconds. The counting time could have been set to 3 seconds or 7 seconds. The important thing is that both the source and no-source spectra, along with the traditional method, are tested using the same time interval. The actual reason 5 seconds is used is because previous research, (Meengs 2018) and (Brogan 2018), used the same time interval.

The idea is to test for any coincidences between different energy bins. For example, one might expect that there might be a correlation between a bin containing the photopeak and a bin containing the Compton edge whenever there is a signal of interest for a given radionuclide. The correlation would be expected to be absent for a purely background measurement. The regression coefficients β_0 and β_1 will be tested for their efficacy as decision parameters. The priors of this model play a minor role since it is preferred that the model be led solely by the counts provided. In all cases, the prior means μ_0 and μ_1 will be set to zero. Recall that in

Bayesian statistics the prior actually biases the estimate, the strength of this bias being determined by one's level of confidence in previous data. This is not always desirable, especially in this case where we do not seek to bias our estimate on a previous result.

Methods I – Full conditionals

In Bayesian analysis, it is typically desirable to express the posterior distribution as a conjugate to the prior distribution (the prior thus being called a "conjugate prior"). This property, called conjugacy, occurs when the posterior distribution type is the same as the prior distribution type (e.g., they are both Poisson or both normal). A formal definition of a conjugate prior from Hoff (2009) is provided:

A class $\mathcal{P}$ of prior distributions for θ is called conjugate for a sampling model $p(y|\theta)$ if

$$p(\theta) \in \mathcal{P} \Rightarrow p(\theta|y) \in \mathcal{P}.$$

This property makes calculation of posteriors very easy because it can be expressed as the same distribution as the prior whose parameters are a function of the likelihood function and prior distribution. Unfortunately, a linear regression is a multi-variable system and thus the posterior distribution is expressed as the joint probability $p(\beta_0, \beta_1)$. The conjugate prior for this expression does not exist. However, recall that since $p(\beta_0, \beta_1|y) = p(\beta_0|\beta_1, y)p(\beta_1|y) = p(\beta_1|\beta_0, y)p(\beta_0|y)$, it is possible to express the posterior of one of the parameters as a conjugate as a conditional probability on the other parameter, assuming the posterior of the other parameter is known. Given the general form of the linear regression given above, we can calculate the conditional posterior of each parameter as a conjugate prior. The posterior of the β_0 parameter given in terms of the other posterior is:

$$\beta_0|\beta_1, y, x, \sigma \sim N(M_0, V_0)$$

$$V_0 = \frac{1}{n\sigma^{-2} + \tau_0^{-2}}$$

$$M_0 = V_0(\sigma^{-2} \sum (y_i - \beta_1 x_i) + \tau_0^{-2} \mu_0)$$

$$\beta_0 = N(M_0, V_0).$$

Likewise, the posterior for the parameter β_1 is given by:

$$\beta_1 | \beta_0, y, x, \sigma \sim N(M_1, V_1)$$

$$V_1 = \frac{1}{\sigma^{-2} \sum x^2 + \tau_1^{-2}}$$

$$M_1 = V_1(\sigma^{-2} \sum (y_i - \beta_0) x_i + \tau_1^{-2} \mu_1)$$

$$\beta_1 = N(M_1, V_1).$$

The parameters μ_0 and μ_1 are the priors of β_0 and β_1 respectively, and τ_0 and τ_1 are the priors for the standard deviation of β_0 and β_1, respectively. The derivation of these posteriors is beyond the scope of this paper but can be found in a standard introductory text on Bayesian statistics, c.f. (Hoff, 2009). When the posterior distributions are expressed as marginal probabilities of the other parameters, they are called full conditionals. These are not technically conjugate priors, but with the use of iterative methods the posteriors can be approximated.

Methods II – Monte Carlo simulation

In situations where obtaining a conjugate prior is not possible, other methods are needed to obtain the posterior distribution. In cases such as this, when we have only the full conditional expressions of the parameters, Monte Carlo simulation is a useful way to look at the posterior distribution.

The Monte Carlo technique used in this research takes advantage of the fact that conjugate priors exist for each parameter conditional on the other parameters. By way of illustration, let the functions for the full conditionals be $f_0 \text{ and } f_1$ for each parameter. The general concept for iterating the posteriors is given by the following loop. Prior to the loop, initial values

must be defined. Recall that these functions are functions of random variables, and thus each iterate is a random variable. The basic idea is shown by the following simplified code.

Set initial conditions

$S \leftarrow (\text{number of iterations})$

$\beta_0^{(0)} \leftarrow (\text{initial value})$

$\beta_1^{(0)} \leftarrow (\text{initial value})$

Run the Monte Carlo simulation of S iterations

for $(s\ in\ 1{:}S)$

$\beta_0^{(s)} \leftarrow f_0\left(\beta_0^{(s-1)}, \beta_1^{(s-1)}\right)$ # Generate values for B0

$\beta_1^{(s)} \leftarrow f_1(\beta_0^{(s)}, \beta_1^{(s-1)})$ # Generate values for B1

$\beta_0[s] \leftarrow \beta_0^{(s)}$ # Store value for B0

$\beta_1[s] \leftarrow \beta_1^{(s)}$ # Store value for B1

end

This type of algorithm is called a Gibbs sampler (Hoff, 2009); it generates a dependent set of the parameters $\beta_0 = \{\beta_0^{(1)}, \beta_0^{(2)}, \ldots, \beta_0^{(S)}\}$ and $\beta_1 = \{\beta_1^{(1)}, \beta_1^{(2)}, \ldots, \beta_1^{(S)}\}$. Notice that the initial values $\beta_0^{(0)}$ and $\beta_1^{(0)}$ are not stored. The loop will gather a total of S iterations. Each iteration will estimate the s iterate of the parameters based on the previous values of both parameters. The values generated in each iterate will be stored in the sets β_0 and β_1 of length S. Each set serves

as an approximation of the posterior distribution. There are some things to keep in mind about these simulations (Hoff, 2009):

(1) They are not in themselves independent models,

(2) They do not generate more information than given in the data or the prior,

(3) They are alternative ways of looking at the posterior distribution.

The number of iterations used depends on how well and how quickly the values "converge" to a particular value. Convergence is not convergence in the strict mathematical sense, but statistical convergence – convergence to a stable mean with a stable standard deviation.

Methods III – Diagnostics

The iterates in the posterior sets β_0 and β_1 should possess the following three attributes:

(1) The long-term behavior of the posteriors β_0 and β_1 is independent of the initial values,

(2) The correlation between consecutive values and previous values is minimized,

(3) β_0 and β_1 have stable and consistent means and standard deviations in a relatively short interval.

Diagnostic tests are used to discern to what degree the generated sets possess these desirable properties. If the generated sets are ill-behaved, the validity of the entire algorithm might be thrown into question. To ensure that the generated posteriors are good approximations, two diagnostic tests are employed: the traceplot and the autocorrelation plot. These plots will help determine the soundness of the generated approximations in addition to how many iterations are needed to obtain a well-behaved approximation.

The traceplot is simply the plot of all the generated points for each iteration. The traceplot given in Figure 7 is an example of a well-behaving posterior with 1000 iterations and initial value of zero.

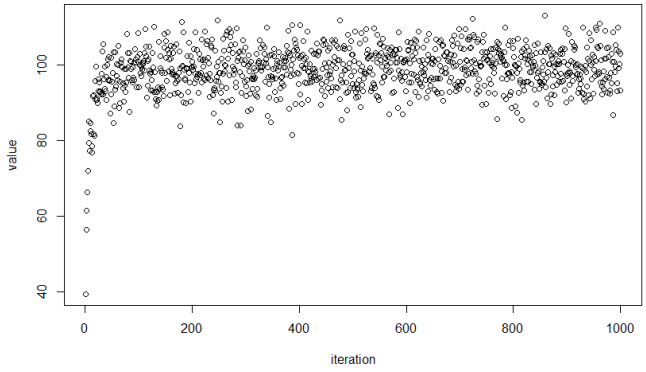

Figure 7 - Example of well-behaving traceplot

It is common practice to jettison the initial values in the posterior to focus only on the well-trained values. In the above plot, the first 100 iterations might by discarded; (in the following models, the first 20 iterations are discarded). An example of a trace plot displaying sub-par attributes is given in Figure 8.

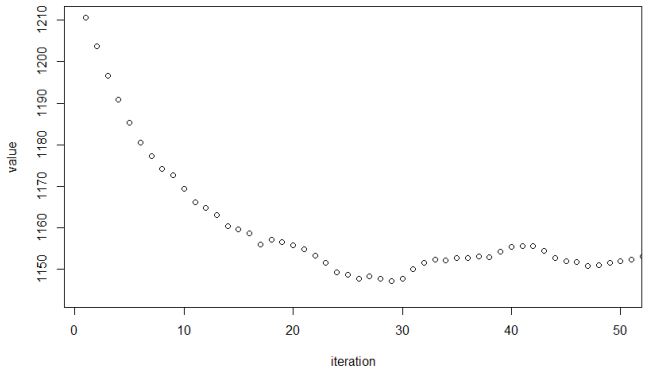

Figure 8 - A traceplot displaying sub-par performance

The plot has only 50 iterations; it is not clear what value the parameter is trying to approach. By increasing the iterations to 1000, we obtain a better-behaved, albeit not perfect, approximation of the posterior. This plot is found in Figure 9.

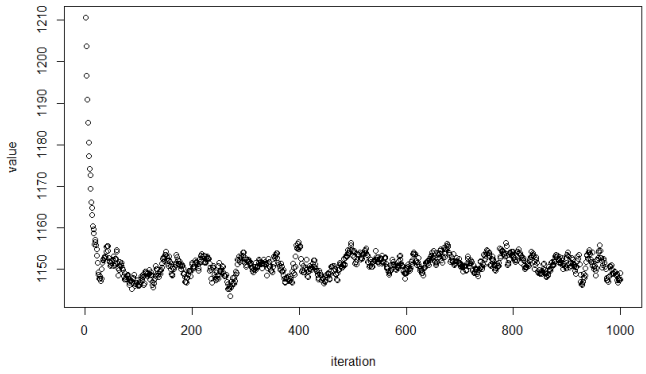

Figure 9 - Traceplot from Figure 8 allowed more iterations

The autocorrelation plot is used to give a visual representation of the correlation of a value to its previous value. The abscissa of the plot is the lag, which measures the “stickiness” of a value to its previous value. The first example given by Figure 10 shows the autocorrelation plot of a well-behaved simulation, especially following the value Lag=60.

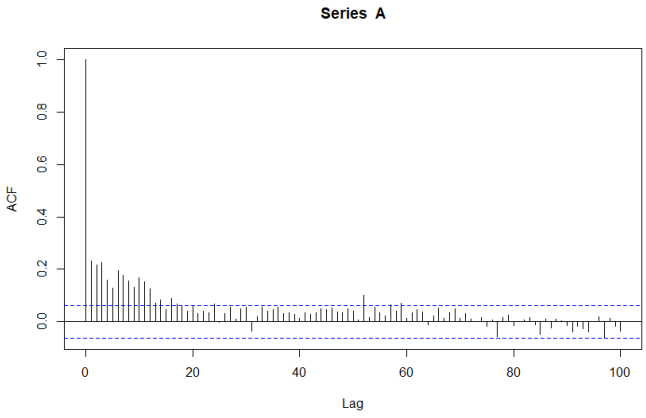

Figure 10 - Example of a well-behaved autocorrelation plot

Unfortunately, with multiple variables the above pattern is not always obtainable. The autocorrelation plot in Figure 11 is a typical autocorrelation plot seen for these algorithms; it displays the tendency to overcorrect, as indicated by the sinusoidal pattern.

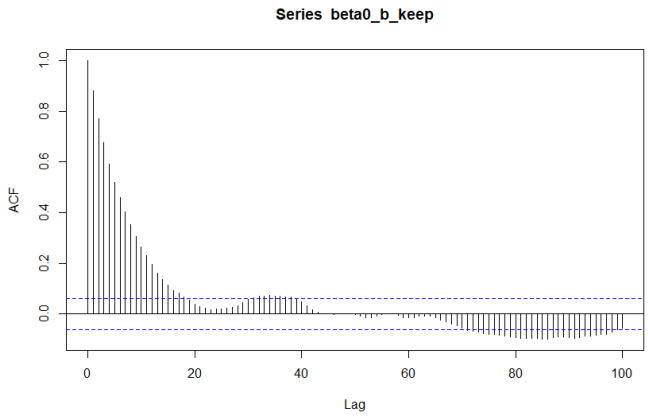

Figure 11 - Autocorrelation plot displaying sub-par characteristics

Autocorrelation is mitigated by simply performing more simulations. If the simulation is repeated 1000 times the sinusoidal fluctuations are "smoothed out." This is shown in Figure 12.

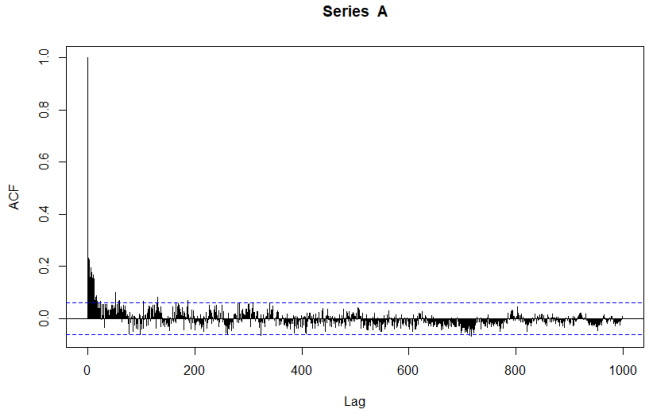

Figure 12 - Autocorrelation plot given more iterations will "average" out the undesirable sinusoidal pattern

After several hundred iterations the sinusoidal behavior somewhat dissipates. This is another reason to jettison earlier iterations, and hence further demonstrating the value of these diagnostic plots.

These two types of diagnostic graphs are examined for each parameter being tested; these graphs are available in Appendix B. The number of iterations S used depends on the parameter being generated. Typically, a value of $S = 100$ is sufficient. There are parameters

that require a larger number of iterations. The sole purpose of these diagnostic plots is to determine the sufficient number of iterations needed to obtain a well-behaved posterior approximation.

Methods IV – The Algorithm

To measure detection efficiency, two sets of spectra are needed: a known background and a known source. The algorithm must be executed for both spectra. Therefore, we obtain two pairs of parameters to be tested against each other: $\beta_{0_{bkgd}}$ vs $\beta_{0_{source}}$, and $\beta_{1_{bkgd}}$ vs $\beta_{1_{source}}$. Additionally, we will examine the β_1 parameters when $\beta_0 = 0$. Thus, for each linear regression, there will be three different tests for detection efficiency (i.e., each parameter will be tested as a decision parameter $\tilde{\beta}$). Each of these tests will proceed in the same manner; so, for the sake of simplicity let β_b and β_s refer to the parameter calculated from the background and the source, respectively. These are the decision parameters. Let β^* be the decision threshold. We apply the criterion $\tilde{\beta} > \beta^*$ to both $\tilde{\beta}_b$ and $\tilde{\beta}_s$ to determine whether we have a false positive and a true positive, respectively. The four possible outcomes are given in Table 5.

Table 5 - The four possible decision outcomes between the source and background

	$\widetilde{\boldsymbol{\beta}}_{\boldsymbol{b}} > \boldsymbol{\beta}^* = \boldsymbol{True}$	$\widetilde{\boldsymbol{\beta}}_{\boldsymbol{b}} > \boldsymbol{\beta}^* = \boldsymbol{False}$
$\widetilde{\boldsymbol{\beta}}_{\boldsymbol{s}} > \boldsymbol{\beta}^* = \boldsymbol{True}$	No Detection	True Detection
$\widetilde{\boldsymbol{\beta}}_{\boldsymbol{s}} > \boldsymbol{\beta}^* = \boldsymbol{False}$	No Detection	No Detection

A true detection is defined to occur only when a detection event occurs for the source but does not occur for the background ($\beta_s > \beta^*$ and $\beta_b \leq \beta^*$). Every true detection will have the effect of perturbing the ROC curve in the concave up direction.

This test must be run multiple times, each with a different sample of the source and background to obtain an understanding of the true detection efficiency for a given β^*. Two new vectors are introduced: a false positive vector for β_b and a true positive vector for β_s. In the simplified code these are expressed as fp and tp. Each iteration will record a 1 or 0 for each true or false positive, respectively. More iterations will bring about finer approximations of the false positive and true positive rates (and subsequently result in smoother ROC curves when β^* is allowed to vary). For instance, if a thousand iterations are used, the approximation of the true and false positive rates will be to the thousandth. The idea is shown by the following simplified code. This code is built around the linear regression algorithm used to calculate both $\tilde{\beta}_b$ and $\tilde{\beta}_s$.

Set number of iterations

$I \leftarrow (\text{Number of iterations})$

Run Simulation

for $(i \; in \; 1{:}\, I)$

*** INSERT LINEAR REGRESSION ALGORITHM LOOP ***

$fp[i] \leftarrow \tilde{\beta}_b > \beta^*$ # Store result of ith iteration for false positive

$tp[i] \leftarrow \tilde{\beta}_s > \beta^*$ # Store result of ith iteration for true positive

end

Obtain final false positive and true positive rates

$fp\_final \leftarrow mean(fp)$

$tp\_final \leftarrow mean(tp)$

R automatically records the results of the logic statement $\tilde{\beta} > \beta^*$ as either a 1 or 0. The false and true positive rates are given by taking the mean of the vectors: $\mathrm{mean}(fp)$ and $\mathrm{mean}(tp)$.

The previous code approximates the false and true positive rates for a given β^* over I iterations. More work needs be done if an ROC curve is to be generated. The decision threshold β^* must vary across a range of values. To do this, another loop is needed (for a total of three loops). A convenient way to do this is to allow β^* to vary between the minimum and maximum $\tilde{\beta}$ value: $\beta^* \in [\min(\tilde{\beta}), \max(\tilde{\beta})]$. Following this last loop, it is a simple task to graph the ROC curve. A simplified code is given below.

Determine minimum and maximum B* values

$\beta_{\min} \leftarrow \min(\beta_b, \beta_s)$

$\beta_{\max} \leftarrow \max(\beta_b, \beta_s)$

Perform loop over all B* values

for (b *in* $\beta_{\min} : \beta_{\max}$)

*** INSERT INNER LOOPS HERE ***

Record positive rates for x and y coordinates of ROC curve

$ROC_x[b] \leftarrow fp\_final$ # false positive rate for b=B*

$ROC_y[b] \leftarrow tp\_final$ # true positive rate for b=B*

end

Graph ROC curve

$\mathrm{plot}(ROC_x, ROC_y)$

This may all seem very tedious; however, the results are promising. The ROC curve is a means, not an end. This simple graphic will immediately impart to us how effective a particular algorithm is at *all* values of β^*, (and therefore at every value for the desired false positive rate p).

These three loops can potentially cause problems with computation time. The total number of iterations is the product of the iterations for each loop. If one is not careful and programs each loop to 1000 iterations (a very modest computation on its own), for example, this will amount to 1000^3 iterations. A billion iterations will usually take many hours for a typical computer. For use in the field, however, only the innermost loop is executed – the loop that is responsible for generating the posterior approximations. The outer two loops are a necessary addition for research: the second loop is needed to compute a true and false positive rate from a long string of "detects" and "no-detects" (i.e., a long string of ones and zeros). The third loop is needed to obtain these true and false positive rates for multiple β^* values. A long computation time due to these loops results merely in an inconvenience for the researcher.

Methods V – Evaluation

It is necessary to compare these results with the traditional method of detection. Other methods of detection have been discussed in the introduction; the traditional method employed in this research is the ISO 11929 (2010) method. This method has also been thoroughly discussed in the introduction. There are several experimental parameters that must remain constant:

(1) Both must use the same data. This ensures that the source, source strength, background, detector, and the geometry of the detector and room are held constant.
(2) The counting time must be constant. This research uses a 5 second counting time for both the traditional method and the new Bayesian methods being tested. Since each spectrum has a 1-second duration, this requires the use of five points of data for each measurement.

A specific procedure is followed when computing these ROC curves for each model. First, a computation is done for both β_0 and β_1. The computation for parameter $\beta_1|\beta_0 = 0$ must be run separately, since β_0 must be set to zero.

Spectra

The pair of background and source spectra used to test these algorithms is given in Figure 13. The noise from low-energy background is quite considerable. The photopeak in the source spectrum is barely discernable at about channel 400. The Compton edge is not discernable in this graphic.

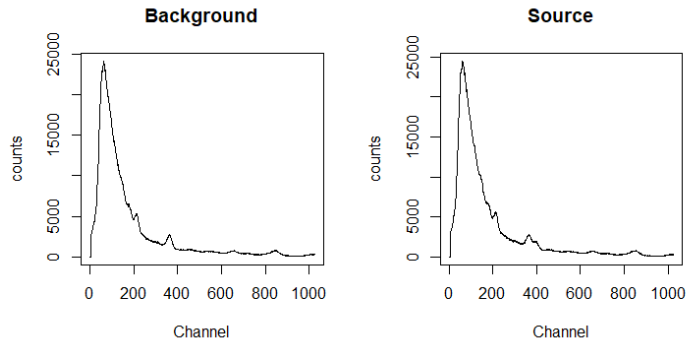

Figure 13 - Background and source spectra used in the Bayesian linear regressions

Results – Testing the eighteen parameters as decision parameters

There are three pairs of Bayesian linear regression models tested. Each pair consists of a gross counts and net counts model. For each of the six different Bayesian linear regression models, three different parameters are examined for a total of eighteen parameters. The eighteen parameters are tested for their effectiveness as a decision parameter. The key idea in these regressions is to compare energy bins that are expected to change in the presence of a source due to physics considerations. It bears repeating that we wish to allow the algorithm to

evolve in accordance with the data; therefore, the weight of the priors will be minimized in all these algorithms. The effectiveness of each of these parameters will be compared to the traditional method of detection. In each case, the ROC curve generated from the traditional method will be plotted using circular points while the ROC curve expressing the Bayesian method will use square points. The two curves are plotted on the same graph. Four different graphs are presented for each parameter, varying by range. The full range is given by the square of the unit interval $[0,1]^2$. The range of each subsequent graph is magnified by a factor of 10 around the origin, (the origin representing the low false positives and true positives; this is our region of greatest interest). The remaining three ranges are $[0,0.1]^2$, $[0,0.01]^2$, and $[0,0.001]^2$. Table 6 provides a summary of the regressions to be tested.

Table 6 - Summary of Bayesian linear regressions to be tested

Equation	Gross Counts	Net Counts
$\boldsymbol{Y_G \sim N(\beta_0 + \beta_1(x_{ce}x_{pp}), \sigma)}$	Regression 1	Regression 2
$\boldsymbol{Y_{pp} \sim N(\beta_0 + \beta_1(x_{ce}), \sigma)}$	Regression 3	Regression 4
$\boldsymbol{Y_G \sim N(\beta_0 + \beta_1(x_{pp} - x_{ce}), \sigma)}$	Regression 5	Regression 6

Regression 1

The first regression utilizes an interaction term as the variable of regression. By taking the product of these two counts, an interaction term is created. This term is compared to the gross counts via a linear regression. The model is given below.

$$Y_G \sim N(\beta_0 + \beta_1(x_{ce}x_{pp}), \sigma)$$

$$\beta_0 \sim N(0,10)$$

$$\beta_1 \sim N(0,10).$$

The independent variable Y_G is the gross counts and the dependent variables x_{ce} and x_{pp} are the Compton edge and photopeak counts, respectively. We now proceed to examine each parameter.

1. β_0 – Intercept term: The results for this parameter are given in Figure 14. This parameter shows little promise of being a useful tool in detecting a source. The $y = x$ nature of this curve indicates that source detection is essentially random.

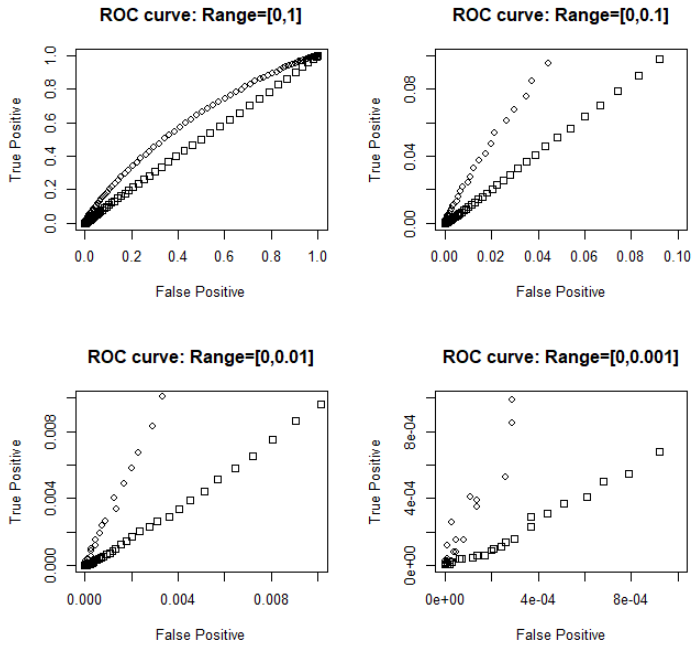

Figure 14 – Regression 1: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_0}$

2. β_1 – Slope term: The results for this parameter are given in Figure 15. This parameter shows an extremely close approximation to the traditional method of detection. In fact, it appears to perform equally well to the traditional method. Additionally, this is an example of where the curve had to be inverted. In other words, the proper detection criterion is given by $\beta_1 < \beta^*$.

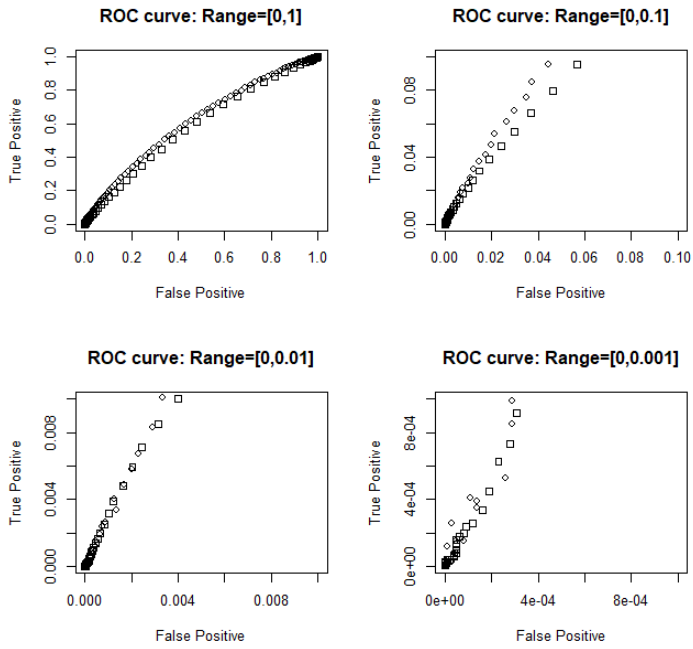

Figure 15 – Regression 1: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_1}$

3. $\beta_1|\beta_0 = 0$ – Slope term through origin: The results for this parameter are given in Figure 16. This parameter seems to slightly underperform compared to the β_1 parameter where the β_0 parameter is allowed to vary. This curve was also inverted, so the proper detection criterion is $\beta_1 < \beta^*$.

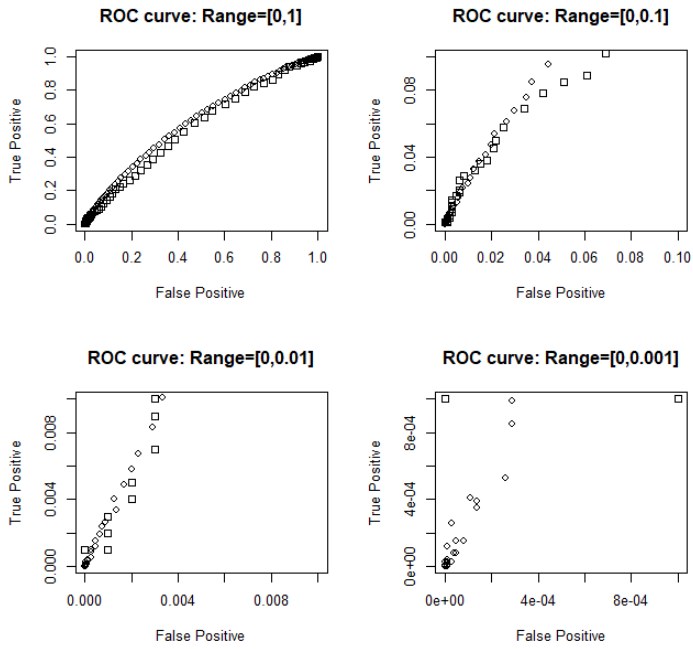

Figure 16 – Regression 1: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_1}|(\boldsymbol{\beta_0} = \mathbf{0})$

Regression 2

As with Regression 1, this regression uses an interaction term between the photopeak counts and the Compton edge counts; except in this case, the net counts are used. Collectively, Regressions 1 and 2 are the interaction-term regressions. In this case, the training background counts from each variable are subtracted from that variable. The model is presented below.

$$Y_G - tr_G \sim N(\beta_0 + \beta_1(x_{ce}x_{pp} - tr_{ce}tr_{pp}), \sigma)$$

$$\beta_0 \sim N(0,10)$$

$$\beta_1 \sim N(0,10).$$

The term tr_G is the gross counts of the training background and tr_{ce} and tr_{pp} are the training background counts of the Compton edge and photopeak respectively.

1. β_0 – Intercept term: The results for this parameter are given in Figure 17. The β_0 parameter for this model performs better than its gross-count counterpart. However, it still underperforms the traditional method.

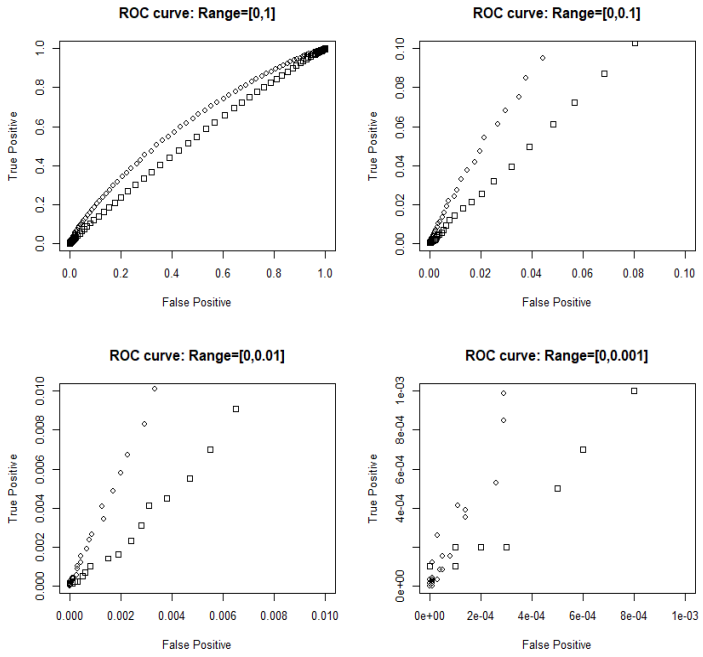

Figure 17 – Regression 2: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_0}$

2. β_1 – Slope term: The results for this parameter are given in Figure 18. This parameter performs as well as its gross-count analogue.

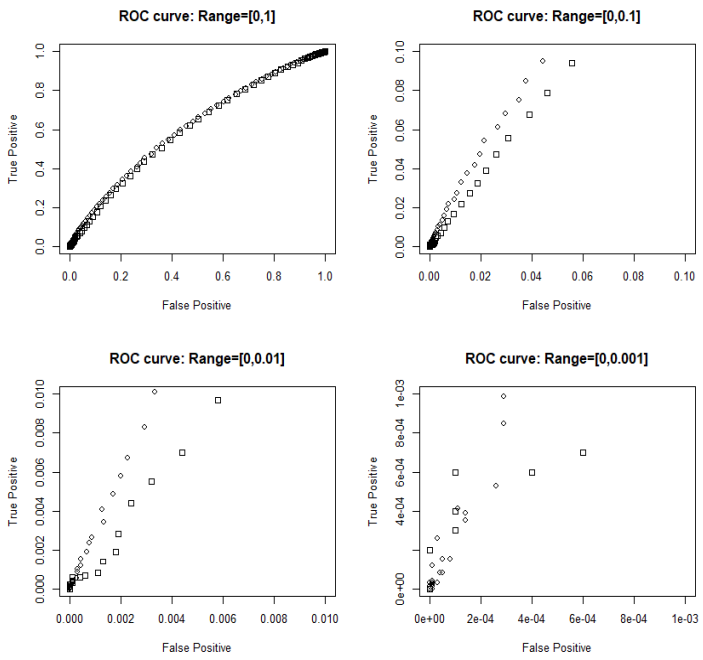

Figure 18 – Regression 2: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_1}$

3. $\beta_1|\beta_0 = 0$ – Slope term through origin: The results for this parameter are given in Figure 19. Although very close, this also underperforms the gross-count version of the interaction model.

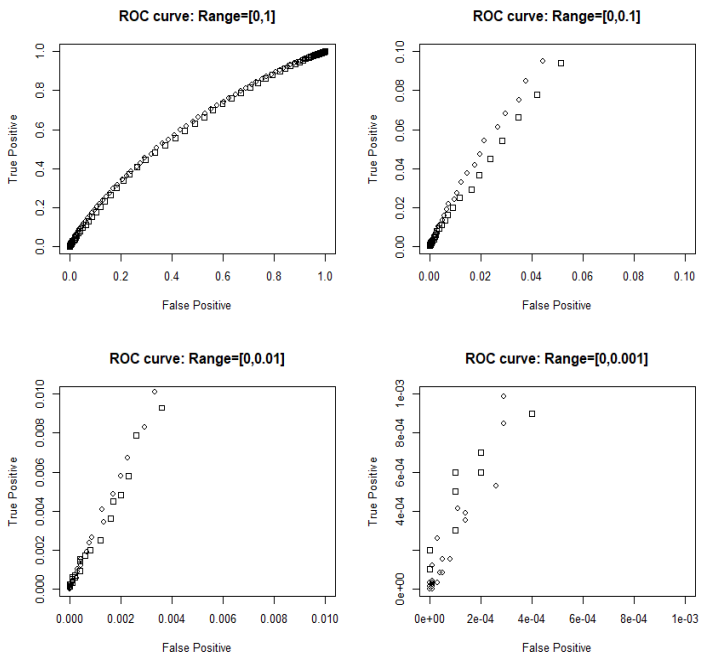

Figure 19 – Regression 2: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_1}|(\boldsymbol{\beta_0} = \mathbf{0})$

Regression 3

In this regression there is no interaction term, but the Compton edge counts are compared directly to the photopeak counts. The model is given below.

$$Y_{pp} \sim N(\beta_0 + \beta_1 x_{ce}, \sigma)$$

$$\beta_0 \sim N(0,10)$$

$$\beta_1 \sim N(0,10).$$

The dependent variable Y_{pp} is the photopeak counts and the independent variable x_{ce} is the Compton edge counts.

1. β_0 – Intercept term: The results for this parameter are given in Figure 20. This is not a well-performing parameter.

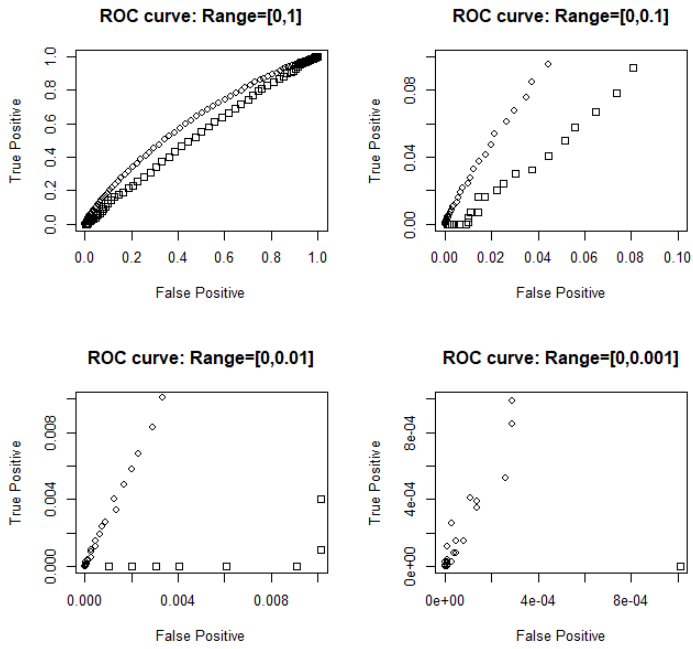

Figure 20 – Regression 3: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_0}$

2. β_1 – Slope term: The results for this parameter are given in Figure 21. This is also not a well-performing parameter.

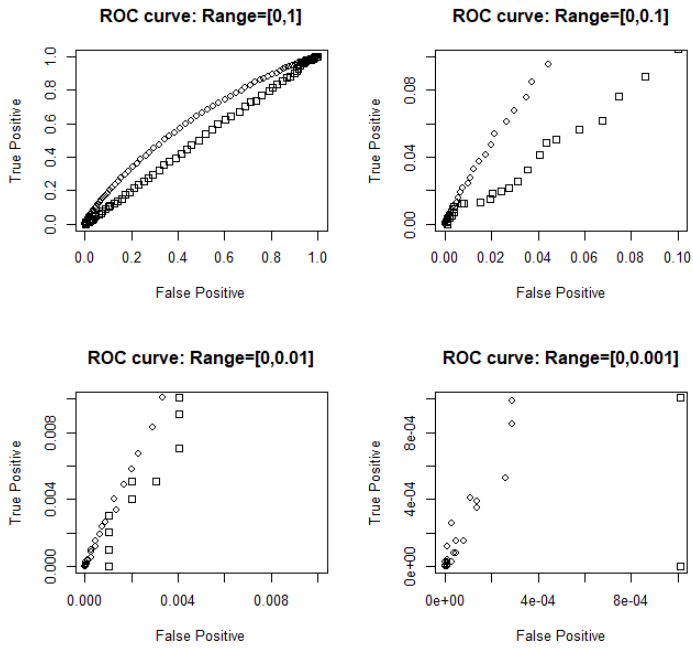

Figure 21 – Regression 3: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_1}$

(3) $\beta_1|\beta_0 = 0$ – Slope term through origin: The results for this parameter are given in Figure 22. This parameter performed very well, slightly less than the traditional method.

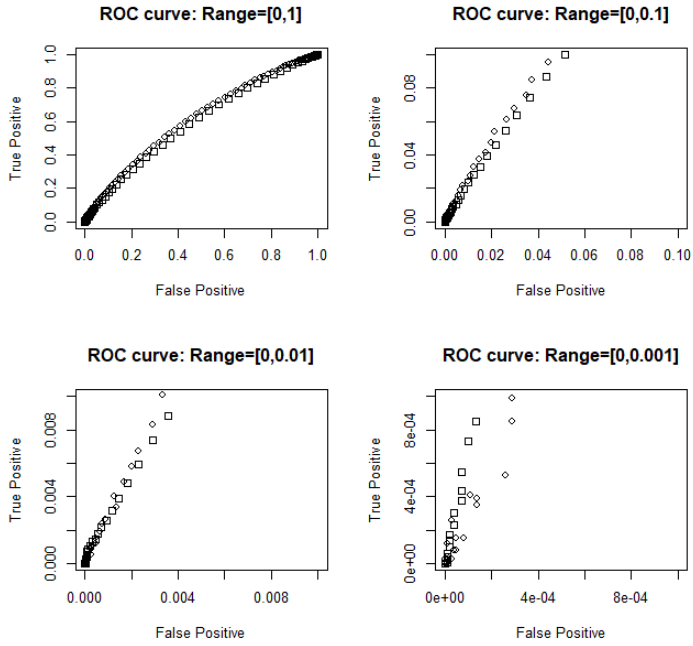

Figure 22 – Regression 3: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_1}|(\boldsymbol{\beta_0} = \mathbf{0})$

The pattern seen at low counts is a very curious anomaly and warrants further discussion.

Regression 4

This regression is the net-count version of Regression 3. Every one of these parameters function very poorly as a decision parameter as evidenced by the "flattened" curve. The model is given below:

$$Y_{pp} - tr_{pp} \sim N(\beta_0 + \beta_1(x_{ce} - tr_{ce}), \sigma)$$

$$\beta_0 \sim N(0,10)$$

$$\beta_1 \sim N(0,10).$$

The variables tr_{pp} and tr_{ce} are the photopeak counts and the Compton edge counts of the training background.

1. β_0 – Intercept term: The results for this parameter are given in Figure 23.

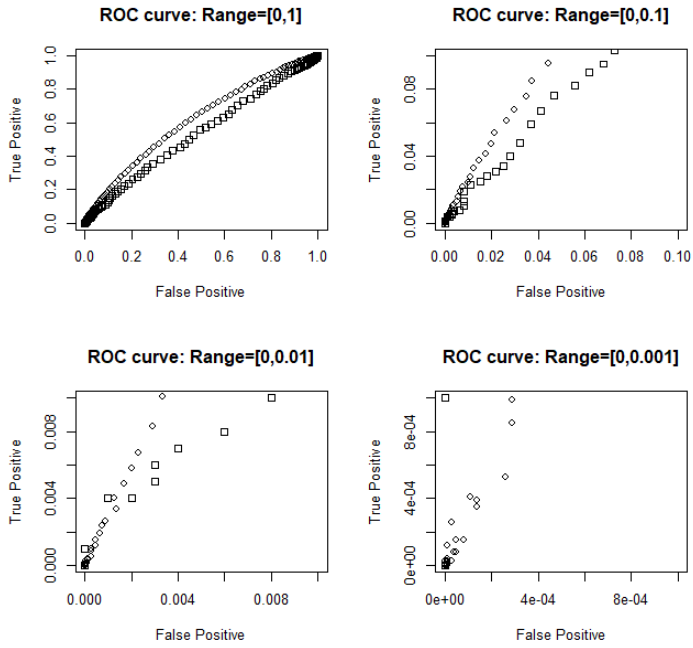

Figure 23 – Regression 4: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_0}$

2. β_1 – Slope term: The results for this parameter are given in Figure 24.

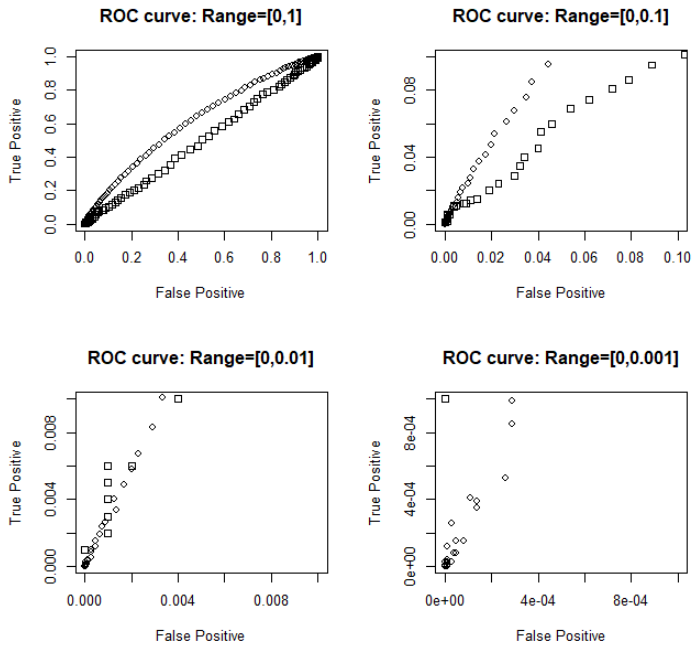

Figure 24 – Regression 4: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_1}$

3. $\beta_1|\beta_0 = 0$ – Slope term through origin: The results for this parameter are given in Figure 25.

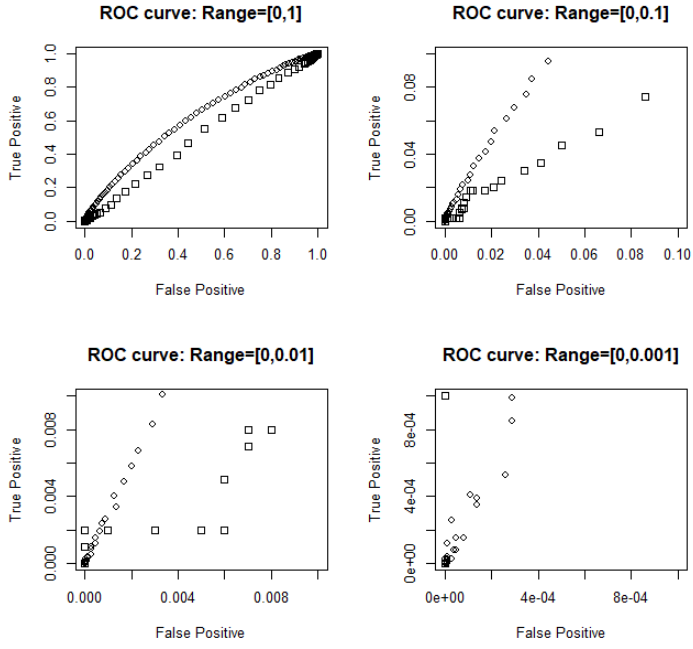

Figure 25 – Regression 4: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_1}|(\boldsymbol{\beta_0} = \mathbf{0})$

Regression 5

The idea behind this regression is to take the difference between the photopeak counts and the Compton edge counts and use that as the variable of regression against the gross counts. Unfortunately, none of the parameters associated with this regression seem to offer any promise. The graphs for these parameters will be presented without further comment. The model is given below:

$$Y_G \sim N(\beta_0 + \beta_1(x_{pp} - x_{ce}), \sigma)$$

$$\beta_0 \sim N(0,10)$$

$$\beta_1 \sim N(0,10).$$

The dependent variable Y_G is the gross counts and the independent variables x_{pp} and x_{ce} are the photopeak counts and the Compton edge counts respectively.

1. β_0 – Intercept term: The results for this parameter are given in Figure 26.

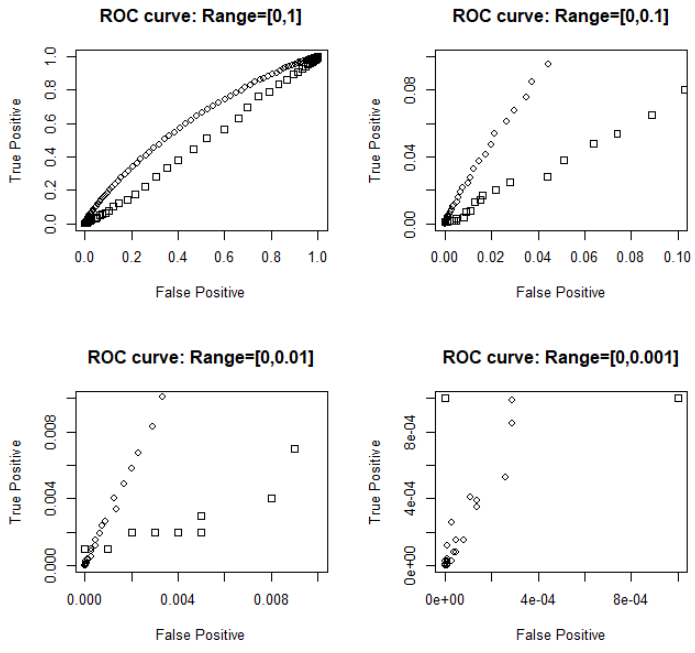

Figure 26 – Regression 5: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_0}$

2. β_1 – Slope term: The results for this parameter are given in Figure 27.

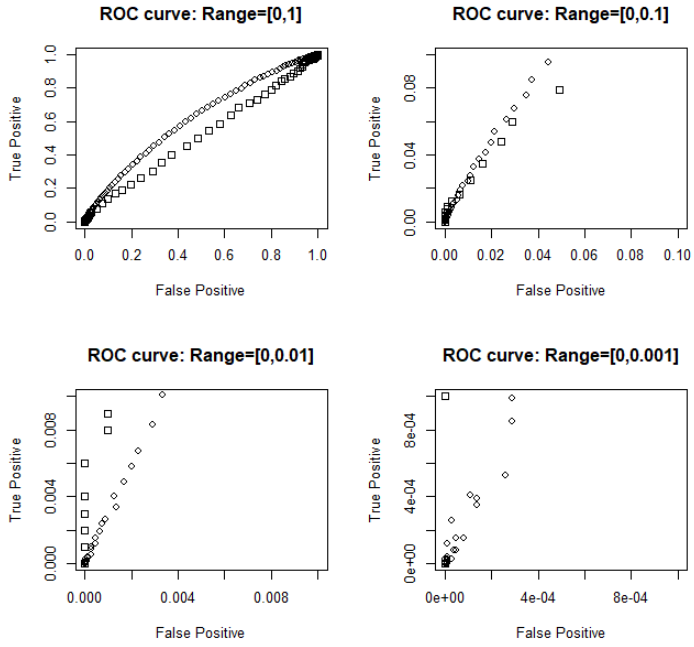

Figure 27 – Regression 5: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_1}$

3. $\beta_1|\beta_0 = 0$ – Slope term through origin: The results for this parameter are given in Figure 28.

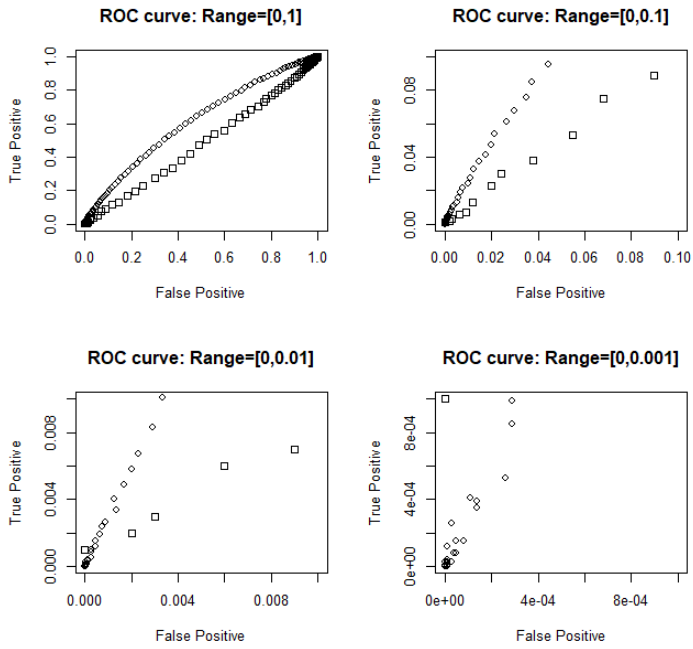

Figure 28 – Regression 5: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_1}|(\boldsymbol{\beta_0} = \mathbf{0})$

Regression 6

This regression is the net-counts analogue to Regression 5. It is presented here.

$$Y_G - tr_G \sim N(\beta_0 + \beta_1((x_{pp} - x_{ce}) - (tr_{pp} - tr_{ce})), \sigma)$$

$$\beta_0 \sim N(0,10)$$

$$\beta_1 \sim N(0,10).$$

The variables tr_G, tr_{pp}, and tr_{ce} are the gross counts, photopeak counts, and the Compton edge counts of the training background respectively.

1. β_0 – Intercept term: The results for this parameter are given in Figure 29. The results of this test are quite surprising. It appears that this exactly matches the traditional method, except at lowers levels of detection, where it seems to perform better. This algorithm was run many times over several hours in order to ascertain this surprising result.

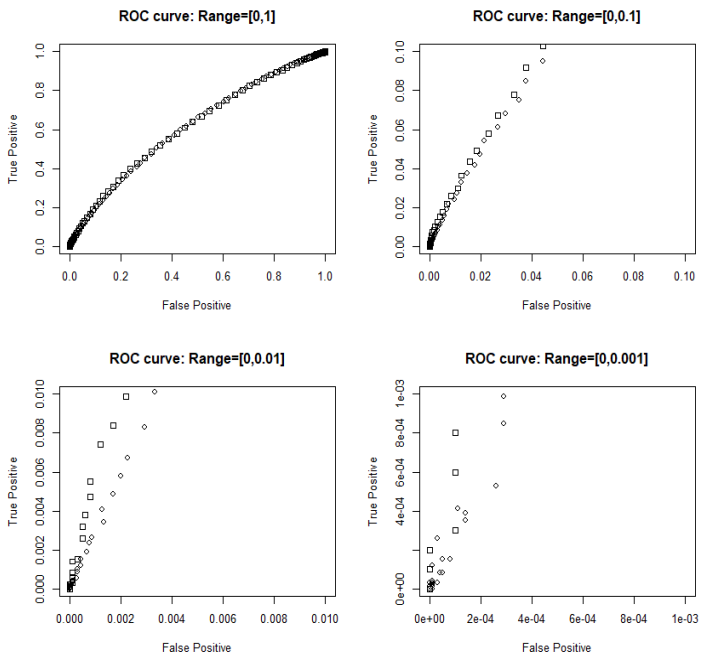

Figure 29 – Regression 6: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_0}$

. β_1 – Slope term: The results for this parameter are given in Figure 30. The linear $y = x$ shape of this curve indicates that this is yet another failed parameter.

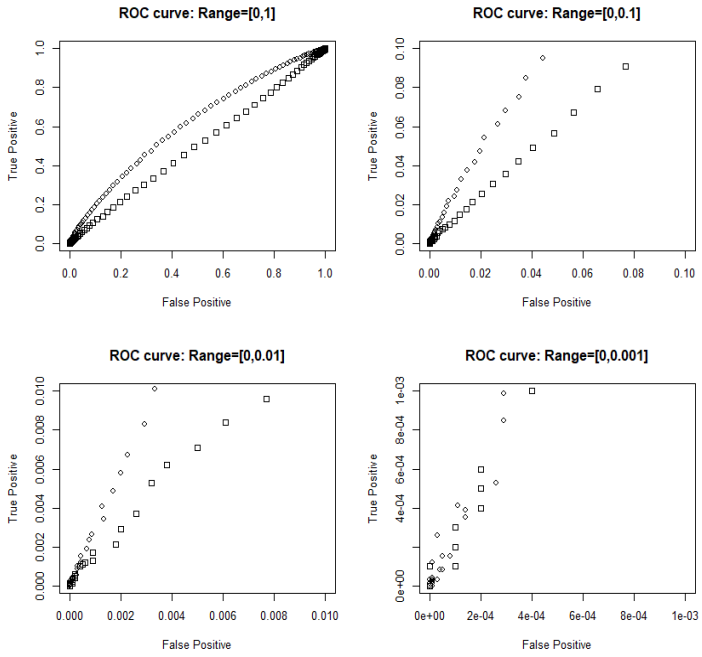

Figure 30 – Regression 6: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_1}$

. $\beta_1|\beta_0 = 0$ – Slope term through origin: The results for this parameter are given in Figure 31.

This is another example of a very poor result – similar to the previous β_1 result.

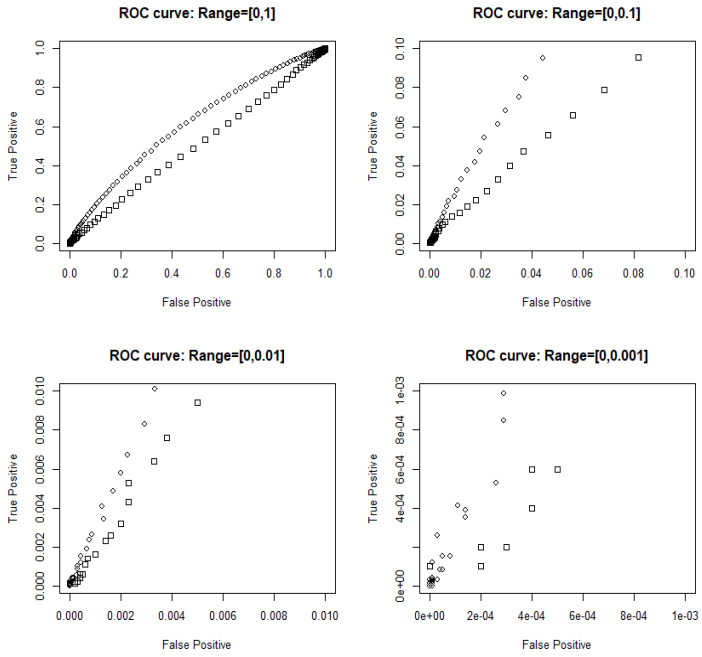

Figure 31 – Regression 6: $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta_1}|(\boldsymbol{\beta_0} = \mathbf{0})$

Discussion

The first regression used the interaction term as the independent variable. The interaction term was the product of the photopeak counts and the Compton edge counts. Only the slope term in Regression 1 was notable. In fact, both slope terms, $\tilde{\beta} = \beta_1$ and $\tilde{\beta} = (\beta_1|\beta_0 = 0)$, perform well as decision parameters. The term where the intercept is forced through the origin performs slightly worse than the term where the intercept is allowed to vary. One interesting thing to notice is that these decision parameters perform nearly as well or as well as the traditional method. As we will see, this is a recurring pattern.

In the second regression, like its gross-counts counterpart, $\tilde{\beta} = \beta_1$ and $\tilde{\beta} = (\beta_1|\beta_0 = 0)$ perform well as decision parameters. The same two parameters, β_1 and $\beta_1|\beta_0 = 0$, perform almost as well as the traditional method; yet they slightly underperform compared to the analogous gross-count parameters. In this case, using the gross counts instead of the net

counts results in better detection. Also, it is not surprising to see that the same two parameters perform well in both models – a trend that is not found in the following models.

In the third regression, the photopeak and the Compton edge are compared directly. Of the three parameters, $\tilde{\beta} = (\beta_1|\beta_0 = 0)$ is the only decision parameter that performs adequately. It performs nearly as well as the traditional method.

There are no parameters that perform well as decision parameters in the fourth regression. They all perform poorly. The pressing question is: Why does the $\beta_1|\beta_0 = 0$ parameter do so poorly in the fourth (net-counts) regression but well for the third (gross-counts) regression?

The fifth regression uses the difference between the photopeak and Compton edge as a parameter. Unfortunately, all three of the parameters tested show no promise of being useful as the ROC curves in the previous section show.

From the previous regression, one might expect that the net-counts version would also be subpar. This is very much not the case. The intercept parameter β_0 of this model is the best parameter tested so far, (the other two not worth pursuing). Again, the results were unexpected as the parameter β_0 performs well with net-counts but poorly for gross-counts (the reverse is true for regressions three and four). This parameter has been shown to perform even slightly better than the traditional method at very low levels of false and true positives. For this reason, the algorithm was run for a long time to obtain the best estimate possible. Additionally, the algorithm was run several times with the same result. Nevertheless, it's still possible that this result could be due to the data used notwithstanding the randomized sampling technique used.

Each of the three sets of models had at least one parameter that shows promise as a detection parameter. Table 7 presents which of these models work best. The qualitative metric of "good" indicates that the parameter performed at par with the traditional method whereas the metric of "very good" indicates that the parameter performed slightly better than the traditional method.

Table 7 - Summary of top performing models and parameters

Model	Parameter	Gross/Net Counts	Performance
$\boldsymbol{Y_G \sim N(\beta_0 + \beta_1(x_{pp}x_{ce}), \sigma)}$	β_1	Gross	Good
$\boldsymbol{Y_{pp} \sim N(\beta_1 x_{ce}, \sigma)}$	$\beta_1 \vert \beta_0 = 0$	Gross	Good
$\boldsymbol{Y_G \sim N(\beta_0 + \beta_1(x_{pp} - x_{ce}), \sigma)}$	β_0	Net	Very Good

One noteworthy thing about these parameters is that they very nearly approximate the traditional method. Even though these slight differences are consistent every time the algorithms are run, they are so minute that they could easily result from systematic errors. This may perhaps be a clue that the traditional method is in fact the optimal detection method and that its detection efficiency can only be approached. The superiority of the traditional method is difficult to prove, however. The results presented here could simply be failed attempts to find the "true" optimal method. Either way, it cannot be unequivocally declared that a novel algorithm has been discovered that outperforms the traditional method of detection. It is not certain that the results are not merely "features" of the routines used to generate these outcomes. This ought to be regarded as a preliminary study. It would be both an interesting and worthy endeavor for other researchers to attempt to duplicate these results with different methods and data.

Training Background, minimizing initial priors, and establishing β^* from background

Even though the training background is not necessary for research, it nevertheless seems appropriate to offer a brief discussion on how to establish the background and the

decision threshold β^* in the field scenario. A Bayesian method will allow for the background to be continuously updated and will essentially be a process of machine learning.

An important thing to remember is that the training background must use the same algorithm as the algorithm being used to test for source presence. The main difference is that the counting time is likely to be much longer. This will result in a parameter with a smaller standard deviation. There are three possible methods that will be discussed about establishing the training background and β^*.

In each example given, the dependent variable is given by y_i and the independent variable is given by x_i.

(1) Establishing Initial Background: Setting priors equal to zero

When establishing an initial background, the prior will be minimized or be made "flat." In this case, the initial values of τ_0^{-2} and μ_0 will be set to zero. Recall that for the algorithm to work, both parameters must be approximated using the Gibbs sampler method. That method will not be repeated here, but the full conditional posterior equations will be provided with some commentary. Remember that in running this algorithm, initial values must be provided.

- For $\beta_0|\beta_1, \sigma$, use the following full conditional:

$$\beta_0 = N(M_0, V_0) \text{ such that}$$

$$V_0 = \frac{1}{n\sigma^{-2}} \text{ and}$$

$$M_0 = V_0\sigma^{-2}\sum(y_i - \beta_1 x_i).$$

- For $\beta_1|\beta_0, \sigma$, use the following full conditional:

$$\beta_1 = N(M_1, V_1) \text{ such that}$$

$$V_1 = \frac{1}{\sigma^{-2}\sum x^2} \text{ and}$$

$$M_1 = V_1(\sigma^{-2}\sum((y_1 - \beta_0)x_i).$$

In both full conditionals, the n is the time over which training background is taken, assuming one-second measurements.

(2) Continuously Updating Background for a growing time window.

If one wishes to continuously update the training background using the previous background as a prior, the following full conditionals will be used. However, this method will indefinitely continue to place an equal weight on past measurements. Because background typically changes, it may be prudent to use this method for an allotted period of time and then refresh background by repeating the procedure given by (1).

- For $\beta_0|\beta_1,\sigma$, use the following full conditional:

$$\beta_0 = N(M_0, V_0) \text{ such that}$$

$$V_0 = \frac{1}{n\sigma^{-2}+\tau_0^{-2}} \text{ and}$$

$$M_0 = V_0(\sigma^{-2}\sum(y_i - \beta_1 x_i) + \tau_0^{-2}\mu_0).$$

- For $\beta_1|\beta_0,\sigma$, use the following full conditional:

$$\beta_1 = N(M_1, V_1) \text{ such that}$$

$$V_1 = \frac{1}{\sigma^{-2}\sum x^2+\tau_1^{-2}} \text{ and}$$

$$M_1 = V_1(\sigma^{-2}\sum((y_i - \beta_0)x_i + \tau_1^{-2}\mu_1).$$

In these cases, the τ_0^{-2} and τ_1^{-2} are the inverse of the variants for the old values of β_0 and β_1, respectively, and the μ_0 and μ_1 are the mean of the old values of β_0 and β_1, respectively.

(3) Continuously updating background for a diminishing importance in past background.

The following method could be used to continuously update the training background despite a changing background. The idea is to introduce a weighting factor $w \in (0,1)$ to apply to the posterior distribution such that the weight of the prior continually diminishes every time the system updates with a new posterior. The factor w would be calibrated in such a way that the algorithm would keep up with a changing background. It would be larger or smaller depending on how well one would want the algorithm to "remember" past measurements. For a rapidly

changing background, w would need to be smaller. The precise value chosen would be based on experimentation. The full conditionals are given below.

- For $\beta_0|\beta_1, \sigma$, use the following full conditional:

$$\beta_0 = N(M_0, V_0) \text{ such that}$$

$$V_0 = \frac{1}{n\sigma^{-2} + w\tau_0^{-2}} \text{ and}$$

$$M_0 = V_0(\sigma^{-2}\sum(y_i - \beta_1 x_i) + w\tau_0^{-2}\mu_0).$$

- For $\beta_1|\beta_0, \sigma$, use the following full conditional:

$$\beta_1 = N(M_1, V_1) \text{ such that}$$

$$V_1 = \frac{1}{\sigma^{-2}\sum x^2 + w\tau_1^{-2}} \text{ and}$$

$$M_1 = V_1(\sigma^{-2}\sum((y_i - \beta_0)x_i + w\tau_1^{-2}\mu_1).$$

Notice that the weighing factor w is placed as a coefficient to every prior term to weaken the bias of the prior. If $w = 0$, the algorithm is identical to the first algorithm "Establishing Initial Background." Conversely, if $w = 1$ the algorithm effectively becomes the second: "Continuously Updating Background." Hence, this third algorithm is a generalization of the first and second.

Deriving the decision threshold β^* from any of these parameters is a simple task with Bayesian statistics. Given a desired false positive rate p, use the quantile function in R. If the decision criterion is $\tilde{\beta} > \beta^*$, the equation for β^* is

$$\beta^* = quantile(\beta, 1 - p),$$

where posterior β for the desired parameter. If the decision criterion is $\tilde{\beta} < \beta^*$, the equation for β^* is

$$\beta^* = quantile(\beta, p).$$

In each of these situations, β is a set of Monte Carlo generated values; (if β took one value, than the quantile() function wouldn't make any sense).

CONCLUSION

This research has focused on improving the detection efficiency of a weak source. The means by which this is attempted are unconventional. Typically, improving detection efficiency is accomplished by longer counting times or using a more efficient detection system. The convenience of a longer counting time does not always present itself. Instead, the focus here was to develop statistical and computational tools to improve detection for a given detector and counting time. Notwithstanding the popularity of Bayesian style statistics, the use of Bayesian modeling in the particular application presented in this research is rather scant in the health physics literature.

In the first part of this two-part research, a (non-Bayesian) optimization algorithm was developed. The purpose of the algorithm was to determine which energy ranges (energy bins) in a gamma spectrum might contribute to a possible detection. This development is a continuation of previous research that studied the efficacy of $\frac{\Delta\mu}{\sigma}$ as a measure of statistical power regarding the change in counts of an energy bin (Meengs 2018). The algorithm was developed and successfully tested: it identifies which energies would contribute to the signal (provided a source other than background is present). This is quite useful if a source is present; however, since a source cannot be presupposed, both scenarios where a source is present and not present must be tested (i.e., a control must be introduced if the science is to be done properly). This inevitably leads to the situation where the algorithm attempts to optimize mere background counts, leading to an increase in false positives. The question is: does this increase in false positives offset any advantage gained from optimization when a source is present? The results indicate yes, it does increase the number of false positives. Therefore, the optimization algorithm is not useful as a means in detecting a possible unknown source. Further ideas are presented that, if implemented, could result in an algorithm that is very effective in the detection of low-level

sources. Further, it might be possible that an algorithm could be used to identify particular sources. These ideas were not explored in this research; instead, they are exciting possibilities of exploration for future researchers.

The second major section of this research is the primary focus of the paper. A total of six different Bayesian linear regression models are tested. Each model includes three parameters that are tested as a decision parameter, making a total of eighteen parameters. Most of these eighteen parameters show little assurance of being useful as decision parameters. Three parameters are singled out as showing the most promise. The three parameters with their corresponding models are presented in Table 7. The third parameter presented has been shown to outperform the traditional method of detection at very high decision thresholds (high β^* – low false positive). This result is so important that the model is re-presented. In the model,

$$Y_G - tr_G \sim N(\beta_0 + \beta_1\left(\left(x_{pp} - x_{ce}\right) - \left(tr_{pp} - tr_{ce}\right)\right), \sigma)$$

$$\beta_0 \sim N(0,10)$$

$$\beta_1 \sim N(0,10),$$

β_0 is used as the decision parameter ($\beta_0 = \tilde{\beta}$ such that $\tilde{\beta} > \beta^*$). This was the best-performing model and yielded positive results. In this model the difference between the photopeak counts and Compton edge counts ($x_{pp} - x_{ce}$) is used as the variable of regression for the gross counts Y_G. The values tr_G, tr_{pp}, and tr_{ce} are the gross counts, photopeak counts, and the Compton edge counts from the training background, respectively. These positive results might simply be due to idiosyncrasies of the data (notwithstanding efforts to mitigate this possibility). The effectiveness of the proposed model closely resembles that of the traditional method. It does indeed outperform the traditional method at low levels, though by a very small amount. The fact that this is commensurate with the traditional method at all values of β^* except for the low values is the source of the skepticism. In fact, all three of the well-performing parameters are nearly identical to the traditional method. In good science, it is imperative to maintain objectivity.

It is therefore recommended that these models (especially the last) be cross-examined by other researchers to ascertain their effectiveness.

It seems also likely that the traditional method is actually a very good method for the detection of a source. This is evidenced by the fact that the best parameters seem to only match its effectiveness. Since the models presented are much more complicated that the traditional method, the new model use cannot be recommended if their effectiveness is merely equal to the traditional method. It is worth mentioning that several parameters have been successfully and confidently eliminated as effective means of source detection. Even if, in the course of time, every parameter presented in this research (even the good ones) are deemed unfit as a detection parameter, it is hoped (at the very least) that this research was fruitful in the elimination of ineffective methods of detection.

Both sections of this research present ideas whose success, though not conclusive, remain possible or even likely. Indeed, it is earnestly desired that the information presented herein will provide insight (perhaps even illumination) to future researchers.

REFERENCES

ANSI N13.30, *Performance Criteria for Radiobioassay,* American National Standards Institute, Health Physics Society; 2017

ANSI N42.23, *Quality Assurance for Radiobioassay Laboratories*, American National Standards Institute, Health Physics Society; 2000

Bevington P, Robinson D. *Data Reduction and Error Analysis for the Physical Sciences 3rd ed.* McGraw-Hill Companies, Inc, 2003

Brogan J. Development of a Decision Threshold for Radiological Source Detection Utilizing Bayesian Techniques Applied to Gross Count Measurements. Colorado State University, ProQuest Dissertations Publishing; 2018. Dissertation.

Brogan J, Brandl A. *Enhancing Test Statistics by Utilizing Data Patterns in Sequential Measurement Strings in Radiation Detection.* Health Physics 115(6), 698-704; 2018

Cember H, Johnson T. *Introduction to Health Physics 4th ed.* The McGraw-Hill Companies, Inc.; 2009.

Cohen J. *Statistical Power Analysis for the Behavioral Sciences 2nd Ed.* Lawrence Erlbaum Associates; 1988.

Currie L. Limits for Qualitative Detection and Quantitative Determination. Anal Chem 40:586-593; 1968.

Fischer JC. Gross count multi-spectra analysis of weak activity sources amid elevated background and various shielding conditions. Colorado State University, ProQuest Dissertations Publishing; 2018. Master's Thesis.

Gelman A, et al. *Bayesian Data Analysis, 3rd ed.* Taylor & Francis Group, LLC.; 2014

Hoff P. *A First Course in Bayesian Statistical Methods.* Springer Science + Business Media, LLC.; 2009

International Organization for Standards (ISO) 11929 1st ed. (2010). *Determination of the characteristic limits (decision threshold, detection limit and limits of the confidence interval) for measurements of ionizing radiation – Fundamentals and application.* Geneva, Switzerland: ISO 11929.

Johnson T. *Radiation Protection: The Essential Guide for Technologists.* Pastime Publications; 2020

Johnson T, Birky B. *Health Physics and Radiological Health 4th ed.* Lippincott Williams & Wilkins; 2012

Klumpp J. Statistical Methods for the Detection and Analysis of Radioactive Sources. Colorado State University, ProQuest Dissertation Publishing; 2014. Dissertation.

Klumpp J, et al. A new approach to counting measurements: addressing the problems with ISO-11929. Nucl. Inst Meth Phy Res A 892: 18-29; 2018

Knoll G.F. *Radiation Detection and Measurement, 4th ed.* John Wiley and Sons, Inc.; 2010.

Lindsay J, et al. Low-Fidelity Spectral Analysis Utilizing a Binomial Discriminator for Weak-Source Detection Decisions. Health Physics 116(5):727-735; 2019.

Luo P, et al. Bayesian Analysis of Time-Interval Data for Environmental Radiation Monitoring. Health Physics. 104(1): 15-25; 2013

Multi-Agency Radiological Laboratory Analytical Protocols Manual (MARLAP). Environmental Protection Agency, Department of Defense, Department of Energy, Department of Homeland Security, Nuclear Regulatory Commission, Food and Drug Administration, US Geological Survey, Nation Institute of Standards and Technology. July 2004

Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) Fig. D9. Environmental Protection Agency, Department of Energy, Department of Defense, Nuclear Regulatory Commission, Rev. 1; August 2000

McElreath R. *Statistical Rethinking.* Boca Raton, FL: CRC Press; 2016.

Meengs M. Detection of a Weak Radiological Source in Ambient Background Using Spectral Analysis. Colorado State University, ProQuest Dissertations Publishing; 2018. Masters Thesis.

Meengs M, et al. Optimization of Spectral String Data Analysis Using a Binomial Discriminator for Weak-Source Detection Decisions. Health Physics. 117(1):28-35; 2019

Miller G. *Probabilistic Interpretation of Data.* Santa Fe, NM; 2013

National Council on Radiation Protection and Measurements*. A Handbook of Radioactivity Measurements Procedures, 2nd ed. NCRP Report Number 58.* Bethesda, MD; 1985

Neyman J, Pearson E.S. *On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I.* Oxford University Press; 1928

Neyman J, Pearson E.S. *On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part II.* Oxford University Press; 1928

Neyman J, Pearson E.S. *On the problem of the most efficient tests of statistical hypothesis.* Philosophical Transaction of the Royal Society of London, Series A 238: 289-337; 1933.

NUREG – 1400, *Air Sampling in the Workplace*, U.S. Nuclear Regulatory Commission, Washington, DC; 1993

Sanquist T.F., et al. *Designing Effective Alarms for Radiation Detection in Homeland Security Screening,* IEEE Transactions on Systems, Man, and Cybernetics, Part C Nov. Vol.38(6), pp.856-860; 2008.

Tandon P, et al. Detection of radioactive sources in urban scenes using Bayesian Aggregation of data from mobile spectrometers. Information Systems 57:195-206; 2016.

The R Foundation, 2018. *The R Project for Statistical Computing.* [Online] Available at: https://www.r-project.org/

Tsoulfanidis N, Landsberger S. *Measurement and Detection of Radiation 4th Ed.* CRC Press; 2015

Turner J. *Atoms, Radiation, and Radiation Protection 3rd ed.,* WILEY-VCH Verlag GmbH & Co.; 2007.

Turner J., et al. *Statistical Methods in Radiation Physics,* WILEY-VCH Verlag GmbH & Co.; 2012.

Wackerly D, et al. *Mathematical Statistics with Applications, 7th ed.* Belmont, CA: Brooks / Cole CENGAGE Learning; 2008.

Wald A. *An Essentially Complete Class of Admissible Decision Functions.* The Annals of Mathematical Statistics, Vol 18, pp. 459-555. Institute of Mathematical Statistics; 1947

APPENDIX A

Insert both sets of spectra

Background ← (insert background spectra)

Test ← (insert spectra to be tested)

Determine number of bins. nrow() function determines number of rows in a

data frame

n_bins ← nrow(*Background*)

Obtain a vector of means for each bin for both spectra

for (b in 1:n_bins)

$\mu_B[b] \leftarrow mean(Background[b,])$

$\mu_T[b] \leftarrow mean(Test[b,])$

end

Calculate $\frac{\Delta\mu}{\sigma}$ for each bin. Assume a previously defined function DMoS() for this

purpose. The DMoS() function is a function of two variables, μ_b and μ_T.

for (b in 1:n_bins)

$D_b[b] \leftarrow DMoS(\mu_B[b], \mu_T[b])$

End

Arrange the D vector in descending order. Record the permutation as the

vector *order* using the order() function.

Also define a vector $index = (1,2, \ldots, n\_bins)$ to record the bin permutations

$index \leftarrow c(1{:}\, n\_bins)$

$order \leftarrow \text{order}(D, \text{decreasing} = \text{TRUE})$

$index \leftarrow index[order]$

Apply the permutation to both μ_B and μ_T.

$\mu_B \leftarrow \mu_B[order]$

$\mu_T \leftarrow \mu_T[order]$

Establish initial conditions

$OP_{Bkgd} \leftarrow \mu_B[1]$

$OP_{Test} \leftarrow \mu_T[1]$

$OP_{Bin}[1] \leftarrow index[1]$

Apply additive criterion. Assume a predefined Boolean function AC() of the two

variables μ_B and μ_T.

for (b in 2:n_bins)

If AC($\mu_B[b]$, $\mu_T[b]$)

$OP_{Bkgd} \leftarrow OP_{Bkgd} + \mu_B[b]$

$OP_{Test} \leftarrow OP_{Test} + \mu_T[b]$

$OP_{Bin}[b] \leftarrow index[b]$

Else

break

end

APPENDIX B

The purpose of this appendix is to provide the trace plots and autocorrelation plots for each of the eighteen parameters. The start of each regression is clearly labeled and is followed by three sets of graphs. The graphs present the generated posterior parameters in the following order: β_0, β_1, and $\beta_1|\beta_0 = 0$. These graphs not only present the behavior of the individual parameters, but the parameters themselves.

Regression 1

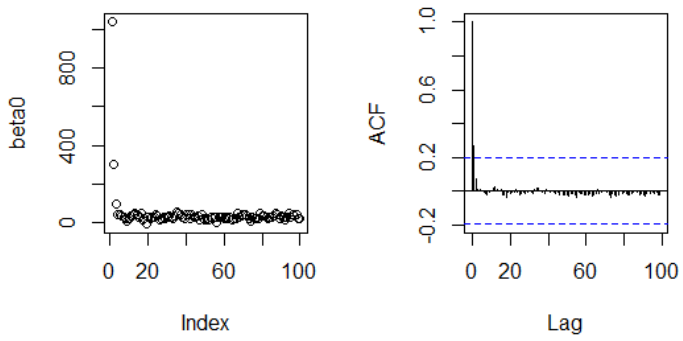

Figure 32

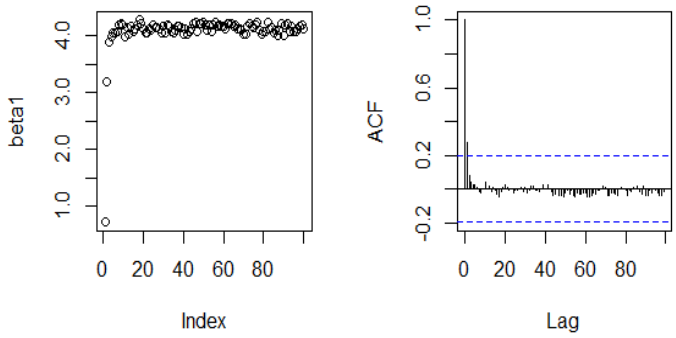

Figure 33

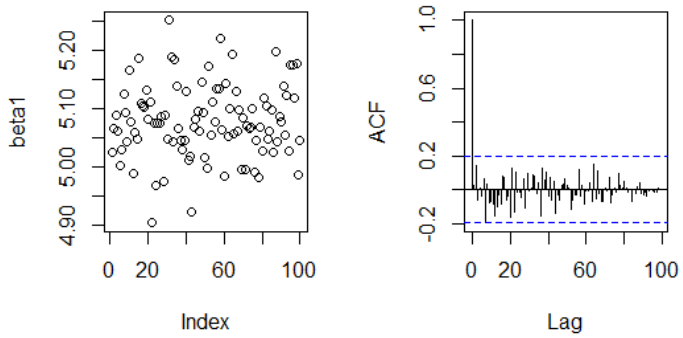

Figure 34

Regression 2

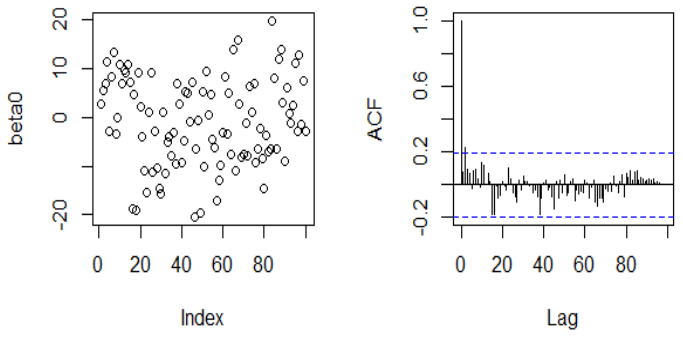

Figure 35

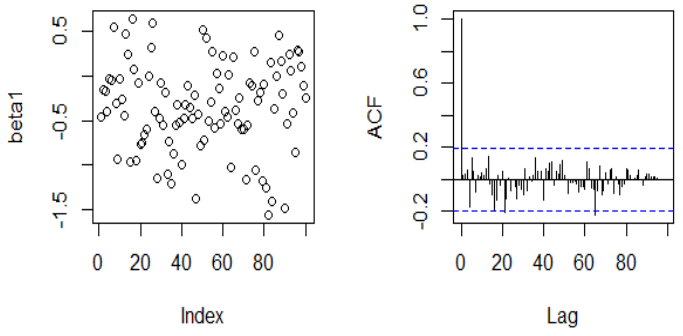

Figure 36

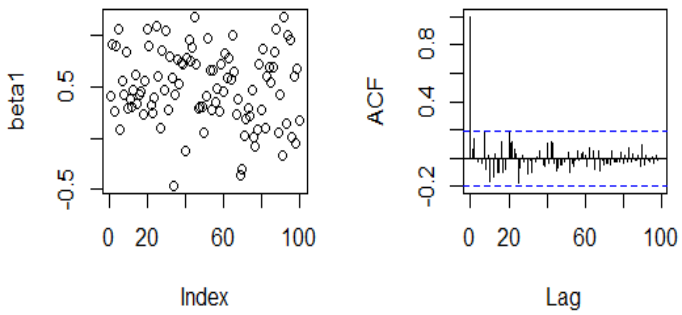

Figure 37

Regression 3

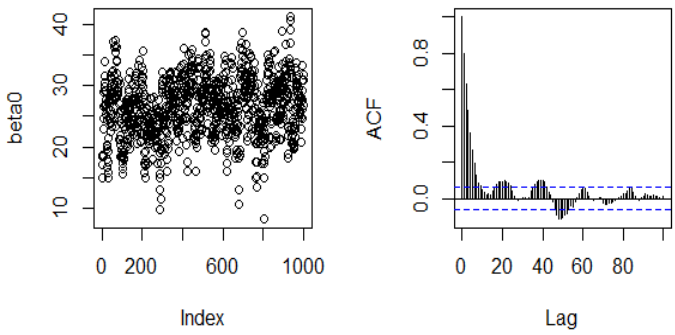

Figure 38

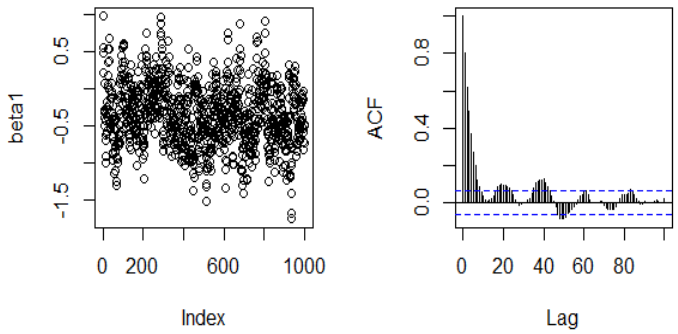

Figure 39

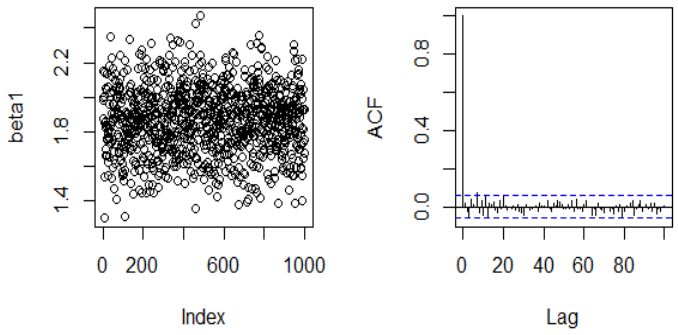

Figure 40

Regression 4

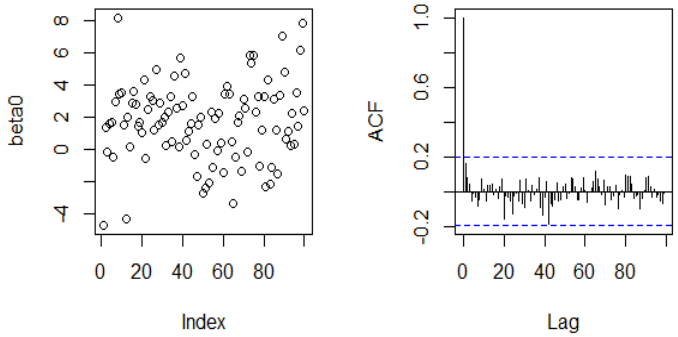

Figure 41

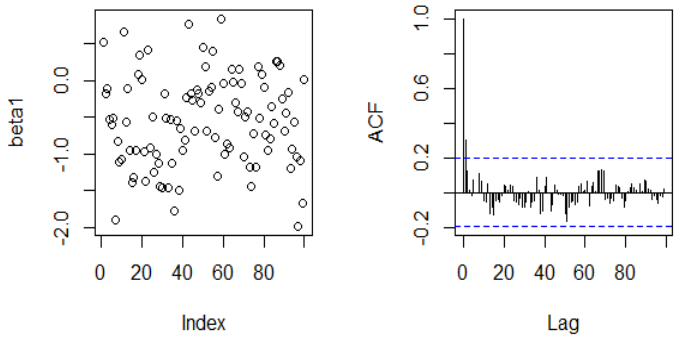

Figure 42

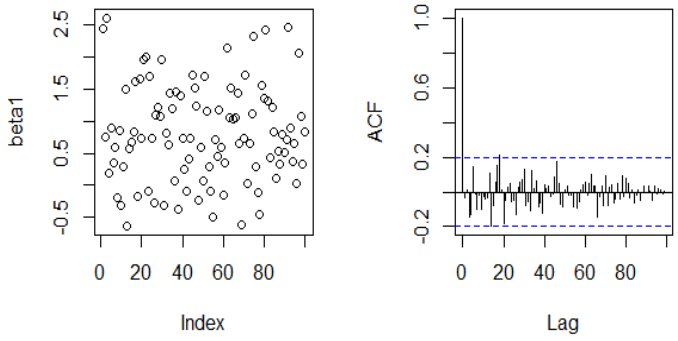

Figure 43

Regression 5

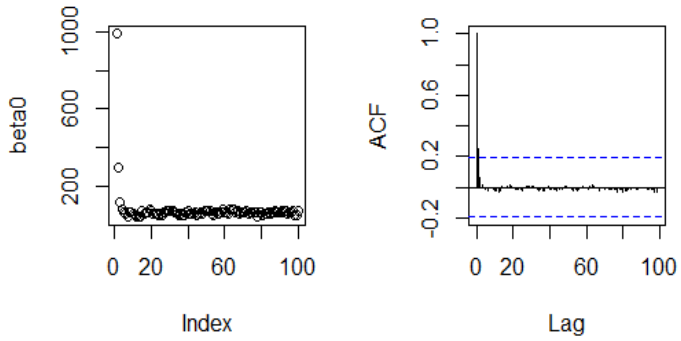

Figure 44

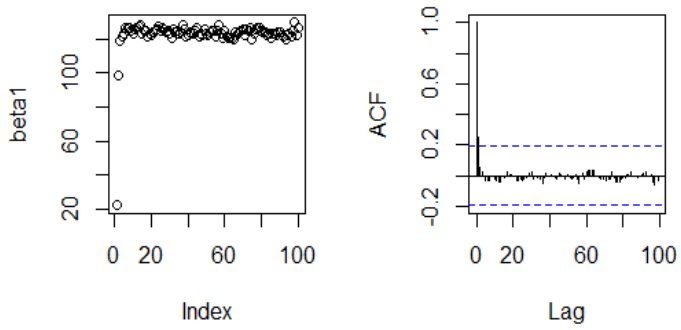

Figure 45

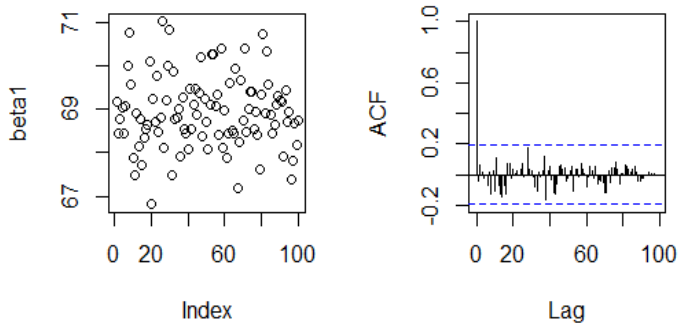

Figure 46

Regression 6

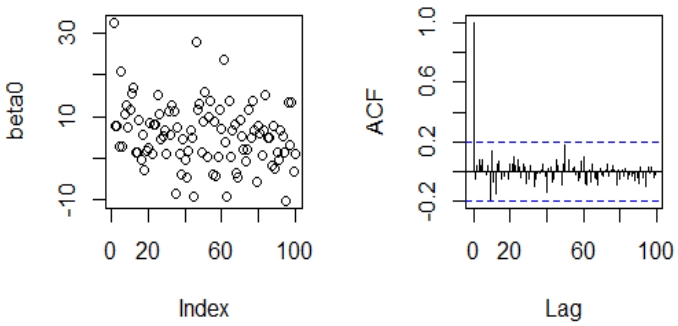

Figure 47

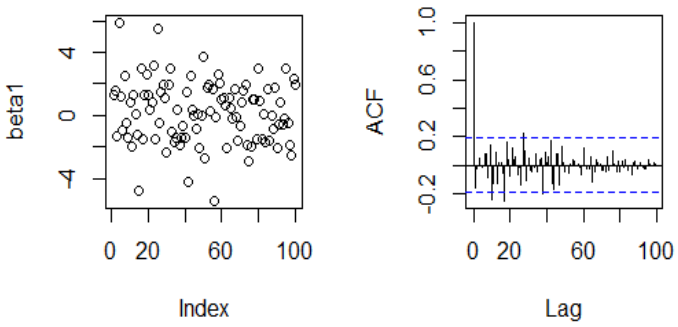

Figure 48

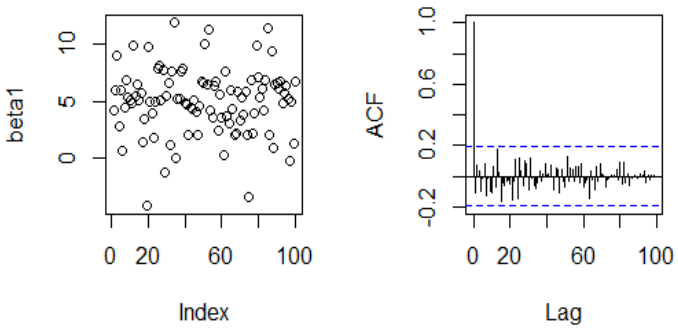

Figure 49

GLOSSARY

The definitions of terms 1 – 6 are provided by Wackerly et al (2008).

1. Population: a large body of data of interest. The population represents the universal set by which all subsets (samples) are obtained. The population and its characteristics cannot be known directly. The population can only be known inferentially through statistical methods.

2. Sample: a subset of the population. Typically, a sample contains more than one value; in this paper, a sample is denoted as a random vector Y. To denote a particular outcome, the vector y is used.

3. Random Sample: a sample in which each member of the population has an equal probability of being sampled. In this research, the random samples are radiation counts within a spectrum.

4. Statistic: a numerical descriptive measure of the sample. A statistic is a function of the data such that $t = u(\mathrm{y})$ where t is a statistic. In a later section, we will introduce the Bayesian linear regressions. The variables x and y will be used as the independent variables and the dependent variables respectively.

5. Parameter: a numerical descriptive measure of the population. A parameter is the value upon which inference is done. A parameter is traditionally denoted by a Greek letter. In this paper, the variable β is used to represent a parameter.

6. Estimator: a rule, usually expressed as a formula, used to calculate the value of an estimate of a parameter based on the measurements contained in a sample. An estimator is denoted as $\hat{\beta}$. An estimator is usually set equal to a statistic: $\hat{\beta} = t = u(y)$. This idea is more commonly used in frequentist statistics.

7. Background: any contribution to the radiation signal not originating from the radiation source of interest. This typically includes naturally occurring environmental radiation such as terrestrial, cosmic, and anthropogenic sources of radiation (Johnson 2020).

8. Source: radiation signal originating from a source of interest. In this research the source of interest is ^{137}Cs.

9. Detection Event: an event in which it is determined that a signal other than background is detected.

10. Decision Parameter: a parameter β that is used to decide if a detection event has occurred. The decision parameter is determined from the data being gathered. In this paper, any parameter β that is used as a decision parameter is denoted by $\tilde{\beta}$.

11. Decision Threshold: a parameter β that is used to decide if a detection event has occurred. The decision threshold is determined from the background measurements. The decision threshold is usually set in accordance with a desired false positive rate. In this paper, any parameter β that is used as a decision threshold is denoted by β^*.

12. Decision criterion: The criterion by which it is determined that a detection event has occurred. The decision criterion is determined by the logic statement $\tilde{\beta} > \beta^*$ (or $\tilde{\beta} < \beta^*$).