DISSERTATION

DEVELOPMENT OF A DECISION THRESHOLD FOR RADIOLOGICAL SOURCE DETECTION UTILIZING BAYESIAN STATISTICAL TECHNIQUES APPLIED TO GROSS COUNT MEASUREMENTS

Submitted by

John Brogan

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

Summer 2018

Doctoral Committee:

Advisor: Alexander Brandl

Thomas E. Johnson
Del Leary
Piotr Kokoszka

ABSTRACT

DEVELOPMENT OF A DECISION THRESHOLD FOR RADIOLOGICAL SOURCE DETECTION UTILIZING BAYESIAN STATISTICAL TECHNIQUES APPLIED TO GROSS COUNT MEASUREMENTS

Numerous studies have been published using Bayesian statistics in source localization and identification, characterization of radioactive samples, and uncertainty analysis; but there is a limited amount of material specific to the development of a decision threshold for simple gross count measurements using Bayesian statistics. Radiation detection in low fidelity systems is customarily accomplished through the measurement of gross counts. Difficulties arise when applying decision techniques to low count rate data, which are restricted by the fact that decisions are being made on individual gross count measurements alone. The investigation presented demonstrates a method to develop a viable Bayesian model to detect radiological sources using gross count measurements in low fidelity systems. An integral component of the research is the process required to validate a Bayesian model both statistically and operationally in Health Physics. The results describe the necessary model development, validation steps, and application to the detection of radiological sources at low signal-to-background ratios by testing the model against laboratory data. The approach may serve as a guideline for a series of requirements to integrate Bayesian modeling (specifically, an interaction model) with radiation detection using gross counts in low fidelity systems.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Alexander Brandl. His mentorship has been invaluable, and I will miss our weekly (sometimes daily) talks. Reflecting on the obstacles we have overcome over the time we have worked together gives me feelings of both accomplishment and comradery, something I consider to only share with my close friends. I look forward to continuing our endeavors.

I would like to thank my mother and father, Jean and Rick, and my brother, Riley, who have been there for me since the beginning of my academic career. I promise no more degrees and graduation ceremonies.

I would like to acknowledge and thank the staff and faculty here at CSU, especially my committee members, Drs. Johnson, Leary, and Kokoszka, for their contributions and guidance on this project. I'm sorry I never went to Japan TJ.

I would like to deeply thank my girlfriend Rachel, for loving, encouraging, and believing in me when I found myself lost or disappointed.

Finally, I would like to continue to express my gratitude to Dr. Hatsumi Nagasawa. I am still on the path you set me on, and my list of achievements continues to grow.

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number, 2014-DN-077-ARI091-01. Disclaimer: The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

TABLE OF CONTENTS

ABSTRACT ii
ACKNOWLEDGEMENTS iii
TABLE OF QUANTITIES AND SYMBOLS vi
EXECUTIVE SUMMARY 1
INTRODUCTION 5
MATERIALS AND METHODS 22
Experimental Laboratory Data Acquisition and Statistical Software 22
Decision Threshold Limitations and Example Data Interpretation 24
Bayesian Linear Regression 30
Bayesian Interaction Model 32
Application to Gross Count Measurement Data 33
RESULTS AND DISCUSSION 38
Format of Investigation 38
Validating the Interaction Model 39
Symmetry in Interactions 43
Using Bayesian Data Analysis to Explore Results 46
Developing γ into a Decision Parameter 48
Testing Under Operationally Equivalent Conditions 52
Model Specification and Optimization 56
Operational Considerations 69
CONCLUSION 75
REFERENCES 80
APPENDIX A 83
Regression and Symmetry Plots Using Bkgd2 as the Training Dataset 83
Regression and Symmetry Plots Using Bkgd3 as the Training Dataset 87
$\gamma diff$ Distributions 90
APPENDIX B 91
$n = 900$ Data 91
$n = 450$ Data 96
$n = 225$ Data 100
$n = 113$ Data 104

$n = 57$ Data 108
$n = 29$ Data 112
$n = 15$ Data 116
$n = 8$ Data 120
$n = 4$ Data 124
APPENDIX C 128
Detection Rates for Various Models for Given $Pr*$ 128

TABLE OF QUANTITIES AND SYMBOLS

θ	The population parameter
θ_0	A specific, accepted value of the parameter θ
$\hat{\theta}$	An estimator of the value of the parameter θ
y^*	Decision threshold established by ISO11929, allows a decision on whether or not a physical effect assessed by the specific quantity to be measured is present
α	**Frequentist:** Probability of type I error in the frequentist hypothesis test; results in a "false-positive" detection when the measurement y is attributed to a measurement of background **Bayesian Linear Regression:** the expected outcome when $\text{predictor}_i = 0$
σ_0	The standard deviation of the net measurement result at zero when the sample contains zero radioactivity
y	Primary measurement result
$u(y)$	Standard uncertainty of the quantity to be measured, associated with the primary measurement result y
L_c	net signal (counts) or result that must be exceeded before there is a specific degree of confidence that the sample contains radioactive material (above background or of the blank)
k_α	Calculated from the cumulative normal distribution, and is the abscissa of the standard normal distribution of the background
L_d	the smallest quantity of radioactive material that can be detected (distinguished from background) with some specified degree of confidence
k_β	Calculated from the cumulative normal distribution, and is the abscissa of the standard normal distribution of the sample distribution
σ_d	standard deviation of the net measurement result when the sample contains radioactivity at the level of the L_d
$\tilde{y}$	A future realization or observation of the measurement y
μ	Mean of the likelihood/sampling distribution
σ	Standard deviation of the likelihood/sampling distribution
μ_μ	Mean of the prior distribution of the parameter μ
σ_μ	Standard deviation of the prior distribution of the parameter μ
min_σ	Minimum value of the parameter σ
max_σ	Maximum value of the parameter σ
β_i	Describes the change in the expected outcome when predictor_i changes by 1 unit in the Bayesian linear regression model
μ_β	Mean of the prior distribution of the parameter β
σ_β	Standard deviation of the prior distribution of the parameter β
γ_i	Parameter to describe the relationship when the outcome_i variable and the predictor_1 variable vary as a function of the predictor_2 variable in the interaction model
$Count$	Value of the observed gross count measurement

$SD5$	The standard deviation of the gross counts obtained in the current measurement and the previous four measurements; used as a predictor variable
$Bkgd\_Sample$	Categorical predictor variable that takes on values of either 0 or 1 depending upon the model description
μ_γ	Mean of the posterior marginal distribution of γ
γ_{diff}	The proportion of differences below 0 after calculating the difference between the marginal distributions of γ for each sample from the posterior

EXECUTIVE SUMMARY

The objective of the work presented is to develop a Bayesian equivalent to the frequentist decision threshold for gross count measurements in radiation detection for low fidelity detector systems. This is accomplished by creating a decision parameter γ_{diff} derived from a Bayesian linear interaction model. The developed model and decision parameter are then tested against operationally relevant scenarios.

Low fidelity detector systems typically require short acquisition times coupled with high sample throughput volume. The most challenging scenario for these systems is one in which the signal-to-background ratio is low; or more specifically when the frequency distribution of the gross count measurements when a source is present is similar to the gross count background distribution. The frequentist decision threshold will always be limited by making decisions on a single measurement. The Bayesian linear regression model provides a posterior distribution for the relationship between the observed gross count measurement and the standard deviation of that measurement and the previous four measurements (termed $SD5$). This relationship can be conditionally modeled upon whether or not the measurement data are categorized as a series of background measurements or a series of measurements in which a source is present. Bayesian regression modeling in which the variables possess a conditional association between the outcome and predictor variables is achieved through the use of an interaction model. The interaction term, γ, then provides a numerical value for the relationship between the observed gross count measurement and the standard deviation of that measurement and the previous four measurements.

An integral component of the results presented is the process required to validate a Bayesian model both statistically and operationally in Health Physics. The results and discussion sections demonstrate the method to develop a viable interaction model to detect radiological sources using gross counts with the parameter γ_{diff} with these principles in mind. The general structure of the procedure is:

- Interaction model validation
- Data exploration using Bayesian data analysis
- Production of an equivalent to the frequentist decision threshold using γ
- Application of the decision parameter to operational conditions and comparison to the frequentist decision threshold
- Testing and expansion of the model using established Bayesian statistical methods and tests
- Application of the results to operational considerations in Health Physics

Validation of the interaction model is achieved by comparing a multivariate regression model with and without interaction to study if the relationship between $SD5$ and gross counts in a measurement depends upon whether or not a source is present. The interaction model is also verified through the fact that these types of models are symmetrical and finding that the relationship between whether or not a source is present and gross counts in a measurement is dependent upon $SD5$. Bayesian statistical analyses are used to understand the parameters surrounding the original relationship that $SD5$ and gross counts in a measurement depends upon whether or not a source is present. This relationship is described statistically by the parameter γ; and the perspective, γ_{diff}, answers the question "what is the probability that the relationship between $SD5$ and gross count measurements from background is less than the relationship

between $SD5$ and gross count measurements from a sample?" This question serves as the framework for a decision rule, $\Pr^* < \gamma_{diff_i}$, which is used in detection scenarios and applied to varying forms of the interaction model. The interaction model is compared to a frequentist decision threshold to determine its detection efficiency. The model performs comparably to the 5s frequentist decision threshold for weak sources at lower false positive rates.

Various forms of the model are developed based on information criterion and tested in the same manner. Widely acceptable information criterion (WAIC) testing did not provide a conclusive model for the best predictive efficiency. Model variations in parameter estimates, flipping the categorical predictor, and changing the nature of the relationship from a linear to a power function display similar detection efficiencies. Any further model validation using multilevel models is unnecessary due to the computation time required for the Markov Chain Monte Carlo (MCMC) calculations used to approximate the posterior distributions. These results all suggest that the original linear model is sufficient at this time, and that more rigorous modeling techniques are required for any possible improvement in detection efficiency.

The interaction model operates by examining a set of previously recorded background gross count measurements, the training data Bkgd1, and resulting $SD5$ with a set of unknown sample gross count measurements and resulting $SD5$. These two sets of data make up the arrays for $Count_i$ and $SD5_i$. Included in the data array is $Bkgd\_Sample_i$, such that each index is correctly categorized as a known background measurement or an unknown sample measurement. This setup allows the model to work in a way that is intuitive to the operational measurement technique: known background data and resulting estimates are used to create a relationship that is expected to be consistent across all measurements with no source present, and this relationship is compared to the samples in question to judge if a source is present. The set up for this system

requires simultaneous acquisition and analysis of measurements from background and the sample measurements in question. Data collected and believed to be background serves as the training dataset. This component of the detection system is advantageous in that an established training dataset is not necessary and long run background measurements are not required to establish parameter estimates, unlike the frequentist decision threshold. The interaction model is extremely sensitive to statistical differences in training dataset distributions. This result is extremely important to consider given that all of the tests use data where the source is stationary and continuously within the field of view of the detector. In a scenario where the source is not stationary, the gross count measurement per time interval is roughly a function of distance from the active region of the detector. The interaction model would be sensitive to the changes occurring per unit time interval, especially one in which the source is passing by the detector. A feature such as this in a string of measurements may not be detectable in the 5 s equivalent of the frequentist decision threshold. A final consideration for the interaction model is that γ, and consequently this categorical linear interaction model, is universal. Even though the scope of the paper studies the relationship between $SD5$ and gross counts in a measurement, in theory any predictor can be used in place of $SD5$.

INTRODUCTION

Radiation detection in low fidelity systems is customarily accomplished through the measurement of gross counts. These systems are typically the primary mechanism used in scenarios where interrogation time is short, sample volume and throughput are high, and source localization is deemed unnecessary, such as at border crossings or access points to radiation or sensitive areas through a radiation portal monitor. If an alarm is triggered, more rigorous methods of detection and identification are initiated. The function and effectiveness of such a system of passive interrogation are heavily dependent upon the initial alarm. However, in low signal to background situations, an alarm may not trigger and a source can pass through undetected. Ideally, with limitless resources to install additional monitors and extend the time per interrogation for all samples, detection systems could be compensating to account for this drawback. In reality, the only cost effective solution is to improve the statistical analysis utilized by the detection system.

Statistical analysis of a radiation measurement traditionally relies upon the use of the frequentist (classical) statistical test. The general objective of a statistical test is to investigate a hypothesis concerning the values of one or more population parameters. The experimenter will have a theory, or research hypothesis, concerning the parameter(s) of interest. Classically, the support for the research hypothesis is referred to as the alternative hypothesis. This support is obtained, using sample data as evidence, by showing that the converse of the alternative hypothesis, the null hypothesis, is false. The test ultimately relies on a proof by contradiction, in which support for one theory is achieved through lack of support for its opposing theory. The working parts of the statistical test are the test statistic, a function of the sample measurements

on which the statistical decision will be based, and the rejection region (Wackerly, et al., 2008). The rejection region specifies the values of the test statistic for which the null hypothesis, H_0, is to be rejected in favor of the alternative hypothesis, H_a.

In the case of radiation measurements, a statistical analysis is carried out on the assumption that the calculated background distribution derived from a "paired blank" count (Currie 1968; ANSI 2011) or a series of measurements is accurately representative of the true background distribution. In statistical theory (Wackerly et al. 2008), the assumption is that the true population parameter θ, the true mean of the background count distribution, is correctly described by the calculated parameter, θ_0, the calculated mean of the background count distribution. The goal of the analysis is to then test a sample (an individual measurement, a series of measurements, or an average of those measurements), $\hat{\theta}$, against the accepted parameter θ_0.

Use of the frequentist statistical test for radiation detection and measurement arises from the necessity to separate signal from background. A level or limit must then be chosen such that, if a sample of measurements were to contain a number of counts in a time interval greater than the established level or limit, the sample is deemed radioactive. Conversely, if the measurements are below these values, the sample is considered background. Given the elements of the statistical test listed previously, H_0 is if the sample measurements are from a background distribution (are not radioactive), and H_a is if the sample measurements are radioactive (these are generalizations of the hypotheses). The test statistic is commonly a function of sample measurements, and the rejection region is the gross count(s), count rate, or net count(s) values of the sample above that level or limit. ISO 11929 defines these values as characteristic limits (International Organization for Standardization, 2010): the decision threshold, the detection

limit, and the limits of the confidence interval. The decision threshold, y^*, is considered an investigatory level and does not provide any information regarding detection capabilities; the detection decision is purely based on this value. The decision threshold is derived by fixing the probability of type I error (false positive), α, on the calculated background distribution. Inherently, this determines the rejection region. Multiple terms exist in the field of Health Physics for the decision threshold, and examples of these can be seen in the table below.

Table 1-1. Different Terms Associated with the Decision Threshold

Name	**Term**	**Defined in**
Decision level	DL	(ANSI N13.30 Rev.1, 2011)
Critical level	L_c	(Currie, 1968)
Critical value	L_c	(IUPAC, 1995)
Decision threshold	y^*	(International Organization for Standardization, 2010)
Minimum significant measured activity	None	(Turner, 2007)
Critical value of the response variable	y_c	(MARLAP, 2004)

Historically, the decision level (DL) is considered the lowest possible useable "action level" when determining whether or not there is activity present above background. The user typically compares measurements to the DL after the measurement is made (*a posteriori*). One formula for DL is

$$DL(N_b|\alpha) = k_\alpha\sqrt{2N_b} \tag{1}$$

where k_α is found from the cumulative Normal distribution

$$1 - \alpha = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{k_\alpha} \mathrm{e}^{-y^2/2}\,\mathrm{d}y \tag{2}$$

and N_b is the observed number of background counts (Strom & MacLellan, 2001). $2N_b$ is assumed to be a good estimate of the variance of the net number of counts. α is the accepted type I error and is typically fixed at a value of 0.05 (International Organization for

Standardization, 2010). The standardization of the detection limit concept was aided by Lloyd Currie (Currie, 1968), who helped develop a consistent statistical approach to the determination of limits for qualitative detection and quantitative determination. He coined his decision level the critical level (L_c), which is defined as the "net signal (counts) or result that must be exceeded before there is a specific degree of confidence that the sample contains radioactive material (above background or of the blank)" (Rucker, 2001). The critical level was defined such that

$$L_c = k_\alpha \sigma_0 \tag{3}$$

where k_α is the abscissa of the standard normal distribution for the blank or background distribution at the chosen probability, and σ_0 is the standard deviation of the net measurement result at zero when the sample contains zero radioactivity.

The decision threshold, y^*, is defined in a similar manner to Equations 1 and 2. The decision threshold is derived by fixing the probability of a type I error, α, on the calculated background distribution and is found in terms of the primary measurement result, y, and the associated standard uncertainty, $u(y)$, as:

$$y^* = y + k_\alpha u(y), \tag{4}$$

It is common practice to use net counts for the parameter values in the frequentist hypothesis test; however, the test also can be applied to only background counts in a simplified test. If the background count distribution is well known, the primary measurement result y from Equation 4 becomes the mean of the measured background count distribution, and the uncertainty associated with this result is estimated by the standard deviation of the background count distribution.

Since their first inception in the field, the characteristic limits have been elaborated on numerous times in order to develop practical applications. Unfortunately, this has led to confusing nomenclature and an overlap between detection decisions being made using a

threshold based on type I error and minimum detectable concentrations (MDC) using type II error (false negative). The concept of MDCs was established before Currie's 1968 paper (Altschuler & Pasternack, 1963), but it was Currie's work that helped standardize the definition. Currie characterized the detection limit, L_d, as "the smallest quantity of radioactive material that can be detected (distinguished from background) with some specified degree of confidence." A considerable difference between L_d and L_c is that L_c is determined *a priori*, before the measurement is made. The calculation for L_d is

$$L_d = k_\alpha \sigma_0 + k_\beta \sigma_d \tag{5}$$

where $k_\alpha \sigma_0 = L_c$, k_β is the abscissa of the standard normal distribution for the MDC distribution, and σ_d is the standard deviation of the net measurement result when the sample contains radioactivity at the level of the L_d. The detection limit is commonly used to compare different methods' measurement capabilities and ability to show compliance with regulatory limits. The detection limit is also referred to as the MDC (MARLAP, 2004), detection limit ($y^\#$) (International Organization for Standardization, 2010), minimum detectable true activity (Turner, 2007), minimum detectable amount (MDA) (ANSI N13.30 Rev.1, 2011) and lower limit of detection (LLD) (NUREG-4007, 1984). A generalized equation for the LLD is

$$\text{LLD} = (4.66 \times \sigma_d) + 3 \tag{6}$$

in which both type I and type II errors are fixed at 5% and the sample and background counting times are equal (Cember & Johnson, 2009). The second term in this equation, $+3$, takes into account the probability of observing 0 counts at a probability of 5%. LLD can be extended to taking into account the counting efficiency of a system. This is known as the minimum detectable activity (MDA) (Cember & Johnson, 2009), and can be calculated in the following manner:

$$\mathrm{MDA} = \frac{(4.66 \times \sigma_d) + 3}{K \times t} \tag{7}$$

where K is the factor that includes counter efficiency, the conversion factor for changing transformation rate into Bq if SI units are used, or into Ci if the traditional system of units is used, and chemical yield if chemical extraction is involved in preparation of the sample; and t is the sample counting time and background counting time.

Inappropriate applications of the two types of error to detection decisions can also be attributed to a misunderstanding of to which probability distribution to apply the error. The probability of a type I error depends on the value of the parameter specified in the null hypothesis. This probability can generally be calculated, or at least easily approximated. The probability of a type II error can be calculated only after a specific value of the parameter of interest has been singled out for consideration. The selection of a practically meaningful value for this parameter is often difficult (Wackerly, et al., 2008). Applied to radiation measurements, type II error will be "scenario dependent" and, without any specification to the various types of samples that could be tested, selection of the alternative hypothesis is entirely arbitrary. Type I error, and therefore the decision threshold, is the most important metric to consider in radiation detection using a frequentist statistical analysis.

When applying traditional decision rules to very low count rate data, the data yield higher than expected false positive rates (Strom & MacLellan, 2001). This is due to the intrinsic assumption that the experimenter has a precise estimate of the mean and standard deviation of the background distribution. Sufficient estimates of either one cannot be attained in a low count rate situation because the signal-to-background ratio is so low that detecting distinguishable signal counts from background is difficult. This is especially true in count rate or gross counts analysis. In theory, decreasing the predetermined level of α decreases the number of false

positive detections on the background distribution; however, the rate of positive detection of signal will suffer.

The standard method to compute the decision threshold for counting experiments for ionizing radiation (International Organization for Standardization, 2010) does not specify a routine to measure background. For measurements with expectations of low signal-to-background ratios, a static background distribution may be assessed by conducting periodic long-time background measurements (National Council on Radiation Protection and Measurements, 1985), where the resulting background data are modeled to follow a normal (Gaussian) distribution defined by their mean and standard deviation. These parameters are then fixed and used in a statistical test on subsequent measurements to determine whether or not a source is present. As a result, background data that are accumulated over the course of the ensuing measurement analyses are ignored. This wealth of additional background information available from repeated independent or continuous measurements performed with the same instrument can be used to improve the statistical analysis of a detection system. Operational application of these statistical considerations have been described (Brandl & Herrera Jimenez, 2008) developed (Brandl, 2013), and tested in a laboratory setting (Brogan & Brandl, in press). The statistical data analysis in these studies was developed from applications for quantitative methods to optimize a decision between two exhaustive, mutually exclusive hypotheses: that the signal detected is either due solely to background radiation, or that it is due to a radiological source embedded in the background. The analysis was based on comparisons between data strings acquired by continuously operating instruments screening for illicit radiological sources and expected data patterns in the absence of a radiological source when the data strings reflect random sampling from the background distribution. The pre-determined type I error was

retained in the course of the comparison, such that the occurrence of the whole data string used in the comparison does not exceed the accepted level in the absence of a source. Quantitatively, the data string was analyzed for the probability that the null hypothesis, H_0, is erroneously rejected at a probability of type I error α provided n individual measurements in the data string of length N exceed a threshold level determined by α_{ind} computed for the individual measurements in the data string:

$$P(\text{reject } H_0 | H_0 \text{ is true}) = \binom{N}{n} {\alpha_{ind}}^n (1 - \alpha_{ind})^{N-n} \quad (8)$$

The statistical approach is independent of operational sensor integration time, the time duration of the individual measurements in the data string, provided the integration time is generally fixed during data acquisition and individual measurements are independent from each other. The data string metric is used to improve signal detection at lower false positive rates. The approach takes advantage of finding the p-value, or attained significance level, which is the smallest level of significance α for which the observed data indicate that the null hypothesis should be rejected given a specific test statistic (Wackerly, et al., 2008) while simultaneously using a binomial discriminator. An ROC curve from the experimental study can be seen in Figure 1-1 (Brogan & Brandl, in press).

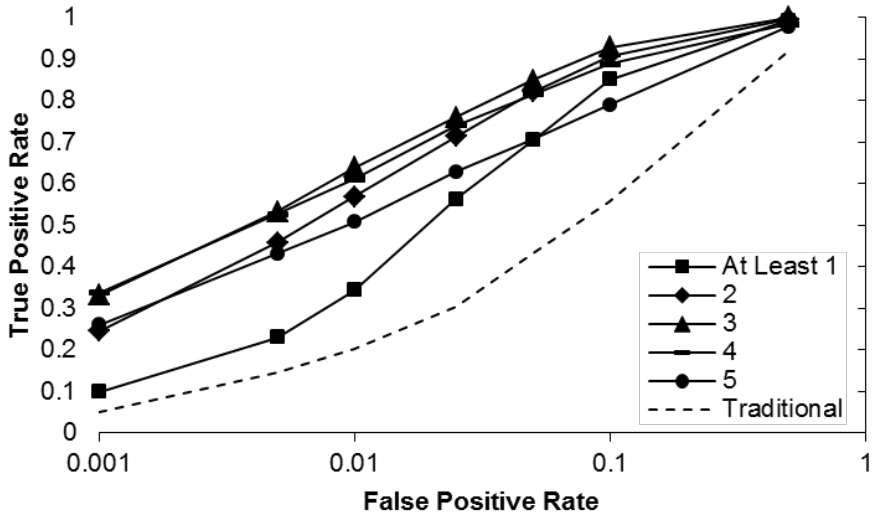

Figure 1-1. ROC Curve Using Data String Algorithm

The figure shows that the traditional decision threshold on gross count measurements can be improved upon using data patterns in a sequence of measurements. This is evident by noting that the Traditional approach (dashed line) had lower true positive rates at a given fixed false positive rate when compared to any of the data string metric approaches in the figure. These studies also validate the assumption that the information contained in prior knowledge can be used to increase detection sensitivity through a "more informed" decision threshold.

Bayesian data analysis, which incorporates the prior beliefs, data, and knowledge of the experimenter and produces probabilities for an underlying parameter, could theoretically be used to create a more informed decision threshold. Moderately similar conclusions result from the Bayesian and frequentist approaches in simple analyses (Gelman, et al., 2013). However, analyses using Bayesian methods are comparatively easier to extend to more complex statistical problems. Numerous studies have been published using Bayesian statistics in source localization/identification (Tandon, et al., 2016), characterization of radioactive samples (Klumpp, et al., 2018), and uncertainty analysis (Michel, 2016); but there is a limited amount of material specific to the development of a decision threshold for simple gross count

measurements using Bayesian statistics. ISO11929 reported a Bayesian approach to the decision threshold using a truncated Gaussian distribution (International Organization for Standardization, 2010), but it was found that extensive revisions were required (Michel, 2016). The difficulty lies in employing the Bayesian interpretation of probability to the characteristic limits defined in ISO11929.

Bayes' theorem allows an investigator to make probability statements about some parameter, θ, given the data, y. The theorem produces a probability distribution of the parameter(s) of interest by combining experimental data with prior knowledge or intuition. This probability distribution, $\mathrm{p}(\theta|y)$, called the posterior distribution, is calculated in the following manner:

$$\mathrm{p}(\theta|y) = \frac{\mathrm{p}(y|\theta)\mathrm{p}(\theta)}{\mathrm{p}(y)}, \tag{9}$$

where $\mathrm{p}(y|\theta)$ is the sampling or data distribution, $\mathrm{p}(\theta)$ is the prior distribution, and $\mathrm{p}(y)$ is a normalizing constant such that $\mathrm{p}(y) = \sum_\theta \mathrm{p}(y|\theta)\mathrm{p}(\theta)$, and the sum is over all possible values of θ (or $\mathrm{p}(y) = \int \mathrm{p}(y|\theta)\mathrm{p}(\theta)d\theta$ in the case of continuous θ). The frequentist decision threshold makes decisions on a probability distribution following the general form $\mathrm{p}(y|\theta)$. In essence, decisions are made on measurement(s) y given the accepted estimate of the true parameter θ using hypothesis testing and confidence interval procedures. The same decision process should not be used on a posterior distribution.

A Bayesian hypothesis cannot be falsified by a significance test, it may only be superseded by competing hypotheses instead (Gelman & Shalizi, 2013). Given this idea, the Bayesian perspective leads directly to a decision theory. The posterior distribution contains all available information and uncertainty about the true value of the parameter θ, and all inference

about the parameter is made on the posterior distribution. In the simplest case, a hypothesis test can be set up in the following manner:

$$\begin{aligned} P(\text{Hypothesis 1 is true}|data) &= \text{posterior probability of Hypothesis 1} \\ P(\text{Hypothesis 2 is true}|data) &= \text{posterior probability of Hypothesis 2} \end{aligned} \tag{10}$$

where the decision criterion is to choose the hypothesis with the higher posterior density. Other established procedures for decision making in Bayesian statistics utilize either a cost function or a Bayes Factor.

The goal of a cost function is to minimize the expected loss for a given hypothesis. The basic loss relationships are a 0/1 loss (L_0), linear loss (L_1), and squared loss (L_2). The calculations for these relationships are

$$\begin{aligned} L_{0,i}(\hat{\theta}) &= \begin{cases} 0 & \text{if } \hat{\theta} = \theta \\ 1 & \text{otherwise} \end{cases} \\ L_0 &= \sum_i L_{0,i}(\hat{\theta}) \end{aligned} \tag{11}$$

$$L_1(\hat{\theta}) = \sum_i \left|\theta - \hat{\theta}\right| \tag{12}$$

$$L_2(\hat{\theta}) = \sum_i \left(\theta - \hat{\theta}\right)^2 \tag{13}$$

The point estimate $\hat{\theta}$ is chosen based on the investigator's choice of loss function, and the value for $\hat{\theta}$ is calculated from the posterior distribution. L_0 is minimized at the mode of the posterior, L_1 is minimized at the median of the posterior, and L_2 is minimized at the mean of the posterior. The optimal hypothesis is that which yields the lowest expected loss.

In the case where two competing hypotheses (H_1, H_2) are being considered, the ratio of prior probabilities of hypotheses can be defined as

$$\text{Prior Odds}(H_1 : H_2) = \frac{\text{p}(H_1)}{\text{p}(H_2)} \tag{14}$$

and the ratio of the posterior probabilities of the hypotheses can be defined in a similar manner, such that

$$\text{Posterior Odds}(H_1 : H_2) = \frac{\text{p}(H_1|data)}{\text{p}(H_2|data)}. \tag{15}$$

Bayes' theorem can be factored into the Posterior Odds formula, giving

$$\text{Posterior Odds}(H_1 : H_2) = \frac{\text{p}(data|H_1)}{\text{p}(data|H_2)} \times \frac{\text{p}(H_1)}{\text{p}(H_2)}, \tag{16}$$

where the first term is called the Bayes Factor (BF). The Bayes Factor quantifies the evidence of data arising from Hypothesis 1 versus Hypothesis 2. In the discrete case, this is the ratio of the likelihoods of the observed data under the two hypotheses. However, in the continuous case, the Bayes Factor becomes

$$BF(H_1 : H_2) = \frac{\int \text{p}(data|\theta, H_1) d\theta}{\int \text{p}(data|\theta, H_2) d\theta} \tag{17}$$

the ratio of the marginal likelihoods. The calculated value for BF is then typically compared to Jeffery's scale, in which ranges of values are associated with varying evidence strength against H_2. Note that if the test was flipped, such that $BF(H_2 : H_1)$, the scale would be associated with varying evidence strength against H_1.

Extending the simple, hypothetical Bayesian hypothesis test to radiation detection is an intuitive process. We can say Hypothesis 1 (H_1) is no source is present, and Hypothesis 2 (H_2) is a source is present. In this simple case, these are the only two possibilities in that they are mutually exclusive hypotheses that cover the entire decision space. A decision must be made between these two hypotheses, and the loss that occurs when decision d is made can be denoted as $L(d)$. The decision space is then (d_1, d_2) where

$$d_1: \text{choose } H_1$$
$$d_2: \text{choose } H_2$$

The Bayesian testing procedure then minimizes the posterior expected loss, which implicitly depends on the relevant consequences. This is typically expressed by the "weight" assigned to the false positive and false negative rates. The investigator then calculates the expected loss for the two different decisions,

$$\begin{aligned} L(d_1) &= P(H_1|\text{source}) \cdot 0 + P(H_2|\text{source}) \cdot \text{weight if } d_1 \text{ is wrong} \\ L(d_2) &= P(H_1|\text{source}) \cdot \text{weight if } d_2 \text{ is wrong} + P(H_2|\text{source}) \cdot 0 \end{aligned} \tag{18}$$

and chooses the decision with the lower expected loss. The entire process heavily depends on the weight selected for a given incorrect decision. Unless specific, verified prior knowledge is applied to the assignment of a loss or weight, use of this decision process is completely arbitrary. It is important to understand that this is not a product of Bayesian statistics; it is a requirement of decision theory (Berger, 1985). According to decision theory, for a given set of information, the optimal action to decide between two options will be to accept one option if and only if the expected posterior loss is smaller than for the other option. The consequences of loss can only be quantified by probabilities; however, the decisive probability for the decision (acceptance and rejection) has to be chosen by humans weighing the importance of loss (Michel, 2016).

A set of decision procedures forming a complete class always includes the optimized decision rule ("complete class theorem"); the set of Bayesian decision rules is such a complete class (Wald, 1947). This implies that any non-Bayesian decision procedure will at least be matched by a corresponding Bayesian procedure, and the "best" Bayesian procedure will provide the optimized statistical analysis. While this theoretically renders Bayesian analyses superior to frequentist approaches, the complete class theorem does not consider possible mathematical or computational difficulties which might be associated with the definition of the optimized

decision procedure; the optimized Bayesian procedure may be mathematically complex or computationally intensive. A Bayesian analysis has to start with a prior distribution which includes the true value of the unknown quantity, so the result must, at least, be known partially (Silver, 2012) (Gelman & Shalizi, 2013), and the investigator must be able to quantify uncertainties inherent in the experiment. As a consequence, selection of a Bayesian process applied to the decision threshold requires an investigation of multiple models. A proper starting point for the development of such models quantifies the distinction between how probabilities are treated in the two statistical disciplines.

The main difference between frequentist and Bayesian statistical approaches may be found in their respective philosophical interpretation and understanding of "probability" (Vallverdu, 2008). In the frequentist approach, probabilities are related to measurable frequencies of different outcomes from observed events. A long-run sequence of repeated events allows for the determination of those frequencies. In the Bayesian approach, probabilities are measures of the state of knowledge or uncertainty of the individual conducting the measurement. It explicitly includes the researcher's knowledge of the experiment and the experimental setup. Considering these two definitions, it can be understood that directly applying the frequentist decision threshold to a Bayesian posterior density is illogical. The frequentist decision threshold utilizes a confidence interval procedure on the probability function $(y|\theta)$, and assumes a proportion of random samples from the same population will produce confidence intervals that contain the true population parameter θ. In the case where $\alpha = 0.05$, we can say that 95% of the random samples will produce confidence intervals that contain the true population parameter. We assume that long repeated measurements of background will yield a valid parameter estimate for the true background. Common misconceptions of the 95% confidence interval procedure are

that there is a 95% chance that the confidence interval includes the true population parameter, and that the true population parameter is in the interval at a rate of 95%. The frequentist definition of probability allows us to define a probability for the confidence interval procedure, but not for a specific sample. The probability that the confidence interval captures the true parameter is either 0 or 1. The problem with this procedure is that we can never know what the probability is that the true parameter will be 0 or 1. The limitation of the confidence interval procedure is then how accurately the true population parameter can be estimated. In radiation detection measurements, we assume that the experimenter has an extremely accurate estimate of the mean and standard deviation of the background. Creating a Bayesian posterior distribution for the background count rate may remediate this complication.

Credible intervals are the Bayesian analogue of the frequentist confidence interval (Kruschke, 2010). They allow the experimenter to describe the unknown true parameter not as a fixed value but with a probability distribution. The construct is similar to the confidence interval, but the key difference is probabilistic statements can be made about the parameter falling within a range of values. Given that a Bayesian posterior distribution is constructed for a background distribution, a credible interval can be used to find how probable it is to observe specific values for the mean of the background distribution. In theory, a summary of the posterior information would provide a sufficient amount of results to start developing Bayesian decision rules for radiation detection once a proper posterior distribution is acquired. "To a Bayesian the best information one can ever have about θ is to know the posterior density" (Chistenson, et al., 2011). One tool that can be used given this distribution is the prediction of future observations $\tilde{y}$. The predictive density of the future observations given the past observations is

$$f_p(\tilde{y}|y) = \int f_p(\tilde{y}|\theta)p(\theta|y)d\theta \quad (19)$$

The same theory that specifies $f(y|\theta)$ also provides a density $f_p(\tilde{y}|\theta)$ for new observations (Gelman, et al., 2013). Given that measurement counts are scalar, the predictive distribution provides the probability that $\tilde{y} \geq \text{count value}$. Theoretically, a "decision" probability can be selected, and the corresponding measurement $\tilde{y}$ would become the Bayesian analog of y^*. The predictive distribution would then provide a Bayesian distribution that functions in the same decision manner used in the frequentist confidence interval procedure: test for observed scalar values that are greater than some limit set by an accepted probability. The strength of a model's predictive efficiency will depend upon the accuracy of the assumptions utilized to construct the model, i.e., the distributions employed for the prior and sampling distributions and the parameter estimates used to describe them.

Specification of the parameters of the priors will always determine the effectiveness of the Bayesian approach (Chipman, et al., 2001), therefore variability in models will hinge upon the various estimators used for these parameters. In the simplest scenario when only conjugate models are applied (Gelman, et al., 2013), the odds ratio test can be used to choose between different models. If parameter choice and model forms become more complex (non-conjugates), such that distributions must be placed on these parameters, hierarchical models must be developed. Computation of the posterior distribution estimates for these models is accomplished using advanced sampling strategies such as the quadratic approximation for MAP estimation and Markov Chain Monte Carlo (MCMC) (Gelman, et al., 2013). Maximum a Posteriori (MAP) estimates are equal to the mode of the posterior distribution, and can be used to generate samples from a parameterized distribution using the MAP estimates to find the form of a posterior distribution. More rigorous model validation techniques must be used if these techniques are

required to estimate the posterior distribution of a model. A commonly accepted test is the Widely Applicable Information Criterion (WAIC) test (McElreath, 2016), which assesses a models' predictive capabilities based on estimating out-of-sample prediction error. The model with the most "weight" scored to it would be the best model to use. It is important to note, however, that WAIC testing and information criterion do not determine what the "best" model is in a group of tested models. They are merely guidelines to help the experimenter gain an idea of a model's predictive capabilities.

In the case of model specification and predictive efficiency, two fundamental types of statistical error leading to poor prediction can occur: overfitting and underfitting (McElreath, 2016). Overfitting results from the model learning too much from the data. Conversely, underfitting is observed when the model does not learn enough from the data. Typically, these two errors are addressed by constructing models using regularizing priors in conjunction with information theory (Gelman, et al., 2013). The goal of regularizing priors is to use estimates that contain enough information to rule out unreasonable parameter values while simultaneously avoiding the possibility of being so strong that the bounds of the estimates rule out values that might be logical to include. Essentially, models must balance between being weakly informative and fully informative. This line of thinking ties into information theory, which dictates proper distribution selection by the principle of maximum entropy and predictive performance measured by information criterion (Gelman, et al., 2013; McElreath, 2016). In essence, distributions selected for models should be the least informative distribution that is still consistent with the experimenter's assumptions. "The distribution that can happen the most ways is also the distribution with the biggest information entropy. The distribution with the biggest entropy is the most conservative distribution that obeys its constraints" (McElreath, 2016).

MATERIALS AND METHODS

Experimental Laboratory Data Acquisition and Statistical Software

The measurement data used in the subsequent analyses presented were collected in a laboratory setup utilizing a 2×2 NaI scintillation detector (Model 902, Canberra Industries Inc., Meriden, CT) with a resolution of 8.5% at the ^{137}Cs photopeak energy of 662 keV. Acquisitions took place at times of the day when the background was expected to be stable. To simulate a low-fidelity system for initial algorithm evaluation and testing, the data were acquired in the open window of the NaI detector system. The data were collected as full spectral data. All channel entries were then summed in the subsequent data analysis to provide simple gross count data. The same laboratory setup was used for data acquisition with and without a source present, representing the presence of an exposed ^{137}Cs source and an evaluation of the background, respectively. Different source "strengths" were simulated by varying the distance between the (point) source and the face of the NaI detector. Measurements were conducted for source-detector distances of 200 cm and 400 cm, and three separate background (Bkgd1, Bkgd2, and Bkgd3) measurements were conducted. Data were collected in measurement sequences to obtain 1800 spectra per background measurement and 1800 source spectra at each source to detector distance. Background and source-detector distance gross count distribution characteristics are summarized in Table 2-1.

Table 2-1. Summary Statistics for Background and Source-Detector Distance Distributions

Statistic	Source-Detector Distance				
	200cm	400cm	Bkgd1	Bkgd2	Bkgd3
Minimum	567	544	540	523	540
Maximum	734	716	706	716	709
Mean	648.7	625.7	617.7	617.3	615.6
Standard Dev.	26.01	26.07	25.37	26.3	26.4

The gross count frequency distributions for the each source-detector distance and background are shown in Figure 2-1. Bkgd1 is overlaid (dashed line) in each figure as a reference.

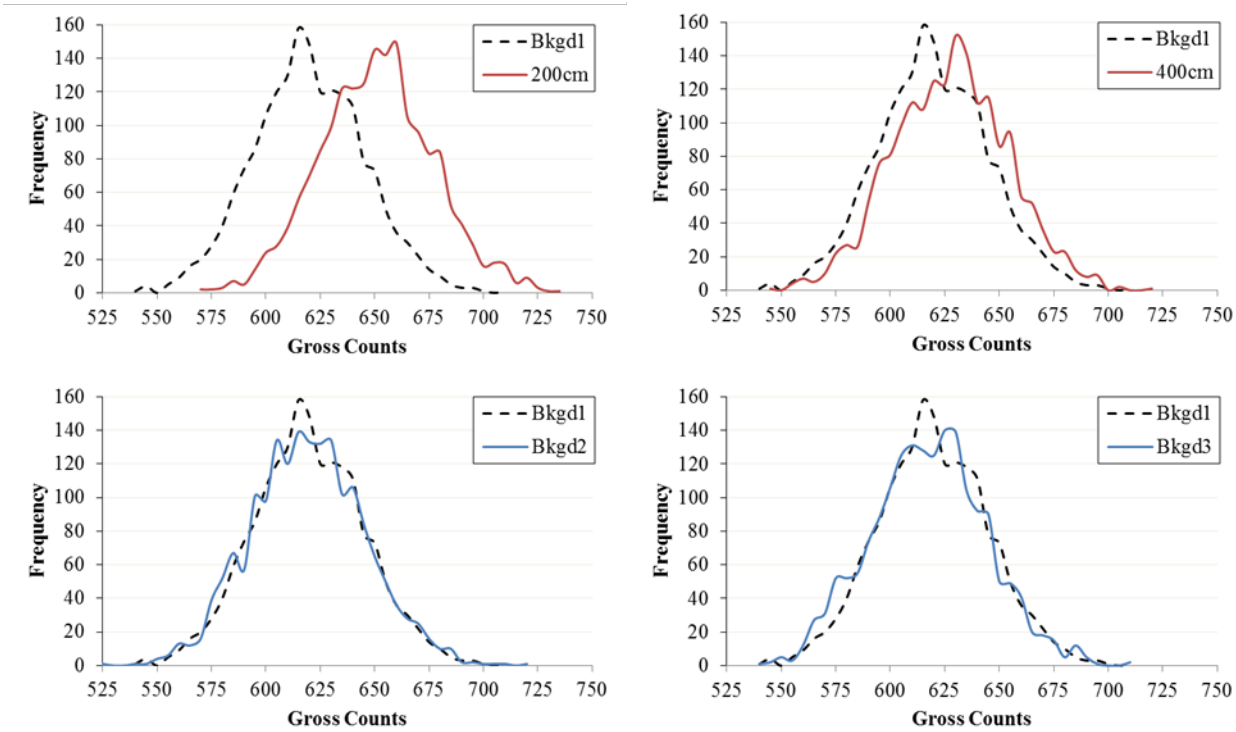

Figure 2-1. Gross count frequency distributions

All statistical tests and analyses performed on these data were done with the R statistical programming language (The R Foundation, 2018) and the Stan probabilistic programming language (NumFOCUS, 2018).

Decision Threshold Limitations and Example Data Interpretation

The frequency distributions in Figure 2-1 display the drawback to making decisions on individual gross (or net) count measurements for low signal-to-background ratio scenarios. Given that the decision threshold, y^*, is calculated using Equation 4 and $\alpha = 0.05$ with the Bkgd1 dataset, $y^* = 659.4$. Using this frequentist decision threshold, approximately 33% of the individual measurements from the 200cm dataset would trigger an alarm and approximately 10% of the individual measurements from the 400cm dataset would trigger an alarm. The limitation is that decisions are being made on individual gross count measurements. As the overlap of the distributions of the background and source increases, the amount of alarms triggered with a source present decreases. A decision threshold on individual measurements will always be limited by this caveat. The same can be said for any Bayesian analogue using a predictive distribution for individual gross count measurements as explained in the previous section. Parameter estimates can be obtained through a Bayesian approach for μ_{Bkgd1} and σ_{Bkgd1}, but the decision threshold will still be applied to individual gross count measurements and the same limitation will exist. However, Bayesian statistics can be used to improve frequentist parameter estimates given specific modeling conditions, or at least come to the same estimate.

The frequentist approach to radiation detection on individual gross count measurements requires that all probabilities be defined by connection to countable events and their frequencies in very large samples. When aiming to accurately model the background distribution in a counting experiment, operationally a health physicist is taught to take a long run count (typically 30 minutes to 1 hour) on a background sample with the goal of estimating the mean and standard deviation of that background distribution to help determine the frequentist decision threshold.

Heuristically, this makes sense, but shows a limitation of using frequentist statistics: frequentist uncertainty is premised on imaginary resampling of data, and as such probability distributions are only applied to the measurements themselves. It requires the health physicist to take a relatively considerable amount of measurements before they can be "confident" in their approximations about parameters such as the mean or standard deviation. The precision of the estimate of a parameter, the standard error (SE), depends upon the number of measurements used in the experiment, further hinging the accuracy of a frequentist approach on the use of very large sample sizes. In contrast, a Bayesian approach to the estimation of the mean of the background allows the experimenter to implement their knowledge into the approximation. All uncertainty can be modeled using posterior probability. Therefore, a model can be built such that probability statements regarding the mean and standard deviation can be made, given the data and the model structure.

A simple experiment was conducted to illustrate the use of both procedures when estimating parameters for the frequentist decision threshold. The Bkgd1, Bkgd2, and 400cm datasets were used for the example data interpretation. n measurements from the Bkgd1 gross count dataset were used to calculate the estimated mean and standard deviation for that background distribution. For example, $n = 3$ corresponds to the first three measurements in the background count data sequence being used to calculate the mean and standard deviation of the background distribution. These parameters would be used to calculate a frequentist decision threshold for a type I error rate of $\alpha = 0.05$ for that n. The resulting decision thresholds per n measurements used were tested against 1800 measurements of the second background data set (to determine the false positive rate) and the source present data set (to determine the true

positive rate). Figure 2-2 shows the true positive rate ($TPrate$) and false positive rate ($FPrate$) on the dataset given n measurements used to determine the frequentist decision threshold.

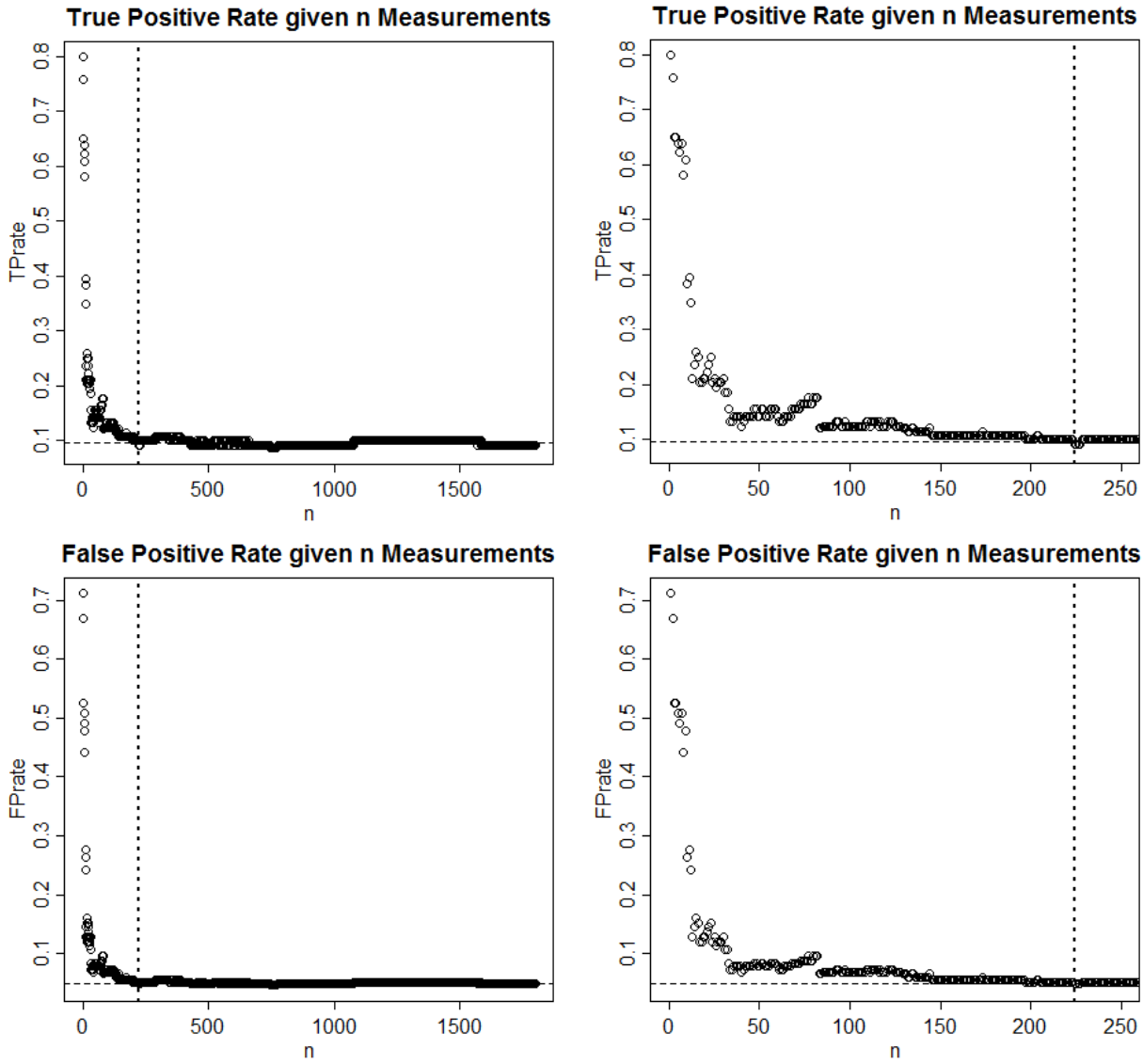

Figure 2-2. True and false positive rates for frequentist decision thresholds determined by n measurements. The left panels are displayed over 1800 measurements and the right panels are displayed up to 250 detailing the onset of the evolution for the decision threshold.

The left graphs show rates up to $n = 1800$ measurements, while the right graphs show the exact same data as the graphs on the left, except n is shown on a smaller scale (up to to $n = 250$). The true positive rate graphs show a horizontal dashed line at $TPrate = 0.095$, which is the true positive rate that the $n = 1800$ measurements decision threshold converges to. The false

positive graphs show a horizontal dashed line at $FPrate = 0.05$, which is the false positive rate that the $n = 1800$ measurements decision threshold converges to. The $n = 1800$ measurement results bear significance because it is the resulting efficiency of a frequentist decision threshold procedure after 30 minutes of background data collected (1800 seconds). The dotted vertical line on all graphs at $n = 224$ displays the first decision threshold that possesses true and false positive rates approximately equal to $n = 1800$. It should be noted that the $TPrate$ did converge to 0.107 by $n = 145$ measurements, but the requirement for the example is to find the exact same $TPrate$ and $FPrate$ as $n = 1800$. To put the dashed line into perspective, approximately the same mean (616.78) and standard deviation (25.56) could have been calculated from a little less than 4 minutes as the parameter estimates from the measurements obtained over 30 minutes. But, a calculation of the 95% confidence interval for the SE of the mean (standard deviation of the measurements divided by the square root of n measurements) demonstrates why the 30 minute parameter estimates are operationally more responsible to use. The 95% SE confidence interval for the 30 minute mean approximation was $[615.58, 617.99]$ while the interval for the 4 minute mean approximation was $[612.80, 620.24]$. This is a relatively wide interval for the mean of a distribution, one that may be unacceptable given the importance of a missed detection or the cost of a false alarm. Regardless, these are intervals of confidence, contingent upon imagined samples and a statistical approach that relies upon the precision of the measurements.

Estimates for the mean and standard deviation of the background can also be obtained using a Bayesian approach. Applying a reasonable likelihood to the gross count data, and prior distributions to the mean and standard deviation of the gross count data, a simple Bayesian model could take the following form:

$$\begin{gathered} Counts_i \sim \text{Normal}(\mu, \sigma) \\ \mu \sim \text{Normal}(\mu_\mu, \sigma_\mu) \\ \sigma \sim \text{Uniform}(min_\sigma, max_\sigma) \end{gathered} \tag{20}$$

The model specifications above state that the likelihood for the gross count data is approximately normal with mean μ and standard deviation σ. The prior distribution for μ is approximately normal with mean μ_μ and standard deviation σ_μ. Lastly, the prior distribution for σ is uniform with some bound on the minimum and maximum values of σ. Applying what is known about the gross count data and some reasonable assumptions, a Bayesian model that would produce a posterior distribution for the mean of the background distribution would take the following form:

$$\begin{gathered} Counts_i \sim \text{Normal}(\mu, \sigma) \\ \mu \sim \text{Normal}(630{,}30) \\ \sigma \sim \text{Uniform}(0{,}50) \end{gathered} \tag{21}$$

This is an acceptable model given what is known about the experiment: the gross count data likely possess a normal distribution, the distribution of μ being normal makes it so the input parameters carry a small amount of information about μ, and the distribution of σ is vague in that equal, positive probability is assigned over a wide range of plausible values. The resulting posterior distribution yields estimates for the values of μ and σ given the data and model — the mean and standard deviation of the background distribution given the measurement data.

Using this model to analyze the data for $n = 1800$ measurements, the posterior approximations of the mean (616.78 ± 0.6) and standard deviation (25.56 ± 0.43) of the measurement data were nearly identical to the frequentist approach. The 95% credible interval, $[615.60, 617.96]$, was also similar to the SE interval of the mean. The difference lies in the interpretation of the values due to the manner in which they were generated. The credible interval allows the experimenter to say: "Given the model assumptions and data, the probability that the mean of the background lies between 615.60 and 617.96 is 95%." For most, this is a

more intuitive statement than: “We are 95% confident that the mean is between 615.58 and 617.99.”

Given what is now known about the posterior estimates for the mean and standard deviation, a more informed model can be built by incorporating these estimates for the mean and standard deviation of the background into the original model. The new model takes the following form:

$$\begin{aligned} Counts_i &\sim \text{Normal}(\mu, \sigma) \\ \mu &\sim \text{Normal}(616.78, 0.6) \\ \sigma &\sim \text{Normal}(25.56, 0.43) \end{aligned} \tag{22}$$

Using this updated model, data for $n = 224$ measurements (equivalent to the 4 minute approximation) were analyzed. The posterior estimate for the mean was 616.75 ± 0.57 and the estimate for the standard deviation was 25.78 ± 0.40. The 95% credible interval, $[615.64, 617.86]$ is much narrower than the frequentist counterpart $[612.80, 620.24]$ at 4 minutes, and is nearly identical to the interval using $n = 1800$ measurements $[615.58, 617.99]$. Herein lays the power of incorporating a Bayesian analysis into parameter estimation. Bayesian modeling allows the experimenter to use prior knowledge to find acceptable estimates with less data and report them in an intuitive manner. To illustrate the impact of prior knowledge on this dataset, the first model (a slightly “non-informed” model) was used to analyze the data for $n = 224$ measurements. The resulting posterior estimates for the mean (616.53 ± 1.85) and standard deviation (27.53 ± 1.31) were comparable to the other models, but the 95% credible interval $[612.90, 620.16]$ fell to bounds similar to the frequentist result on the error of the mean estimate $[612.80, 620.24]$.

This simple example data interpretation displays the differences between using a frequentist or a Bayesian procedure to determine the parameter estimates used to calculate the

decision threshold y^*. Even when a frequentist procedure and Bayesian model give exactly the same numerical answer, the interpretations of these answers due to their statistical nature can have entirely different meanings. Inferences from a frequentist perspective rely upon imaginary, repeated sampling. A Bayesian model treats "randomness" as a property of information, not of the experimental method. Bayesian inference critics often assert that the approach invokes beliefs or subjective opinions, when in reality it is a logical procedure for processing information. Given this line of reasoning, the parameters described in a Bayesian model can be considered targets of learning, which are characterized by the posterior distribution the model produces. In this sense, the Bayesian modeling procedure provides the means to obtain abstract statistical knowledge about the data in question. The most widely applied statistical model to learning from data is a regression analysis. When used in a Bayesian sense, regression analysis provides an intuitive manner to finding the relationship between parameters of interest.

Bayesian Linear Regression

This section utilizes language and terminology for describing and coding statistical models commonly found in statistical texts and journals, and is general to both Bayesian and non-Bayesian modeling.

In a linear regression study, a set of measurements that the experimenter wishes to predict or understand is recognized as the outcome variable or variables, outcome_i. A likelihood function/distribution is defined for each of these outcome variables, providing the probability of observing the specified data under distinct parameter values. These distributions are always assigned normal distributions in linear regression. Separately, the experimenter identifies another set of data, the predictor variables, predictor_i, to use to predict or understand the outcome variables. Linear regression computes the relationship between outcome_i and

predictor_i by linking the shape of the likelihood (i.e., for mean, variance, etc.) to the predictor variables. Under this experiment structure, the experimenter must determine and define all of the parameters for the model. In a Bayesian context, this includes appropriately selecting prior distributions for all of the parameters in the linear regression model. Selection of these priors defines the initial state of the model. The basic structure for the linear regression model is then

$$\begin{gathered}\text{outcome}_i \sim \text{Normal}(\mu_i, \sigma) \\ \mu_i = \alpha + \beta \times \text{predictor}_i \\ \alpha \sim \text{Normal}(\mu_\alpha, \sigma_\alpha) \\ \beta \sim \text{Normal}\left(\mu_\beta, \sigma_\beta\right) \\ \sigma \sim \text{Uniform}(\min_\sigma, \max_\sigma)\end{gathered} \tag{23}$$

where α provides the expected outcome when $\text{predictor}_i = 0$ and the slope, β, describes the change in the expected outcome when predictor_i changes by 1 unit.

The linear model can be expanded to include multiple variables in an analysis that utilizes more than one predictor variable to form an outcome. In multivariate regression, the expected outcome is a sum of at least three independent terms, the intercept α included. A general form for the multivariate regression for two variables is:

$$\begin{gathered}\text{outcome}_i \sim \text{Normal}(\mu_i, \sigma) \\ \mu_i = \alpha + \beta_1 \times \text{predictor}_{1,i} + \beta_2 \times \text{predictor}_{2,i} \\ \alpha \sim \text{Normal}(\mu_\alpha, \sigma_\alpha) \\ \beta_1 \sim \text{Normal}\left(\mu_{\beta_1}, \sigma_{\beta_1}\right) \\ \beta_2 \sim \text{Normal}\left(\mu_{\beta_2}, \sigma_{\beta_2}\right) \\ \sigma \sim \text{Uniform}(\min_\sigma, \max_\sigma)\end{gathered} \tag{24}$$

The multivariate model can be used to study the extent to which an outcome changes as the result of the absence or presence of a category, such as whether an outcome is determined as a background measurement or a set of measurements in which a source could be present. The incorporation of a categorical variable into the above model is achieved by using a predictor

variable that takes on the discrete values 0 or 1, depending upon whether the categorical designation should be absent or present given the data.

Bayesian Interaction Model

The linear models presented up to this point consider each outcome, outcome_i, to be conditional on a set of predictors for each case i. Embedded in these models is the assumption that each predictor has an independent association with the mean of the outcome. Such an assumption is not always correct, as it is conceivable that associations among predictors are conditional (McElreath, 2016). For example, suppose that a relationship between outcome and predictor used in a multivariate model varies according to whether the measurement is from background or with a source present. The previous simple linear models cannot account for the required conditioning. Modeling conditionality where one predictor depends upon another predictor is achieved by using an interaction (Gelman, et al., 2013). A linear interaction model is built by creating a linear model for parameters within the linear model. More specifically, the relationship between outcome_i and predictor_1 is made to vary as a function of predictor_2. Within the linear model, this relationship is measured by the slope parameter β_1. To accomplish the desired interaction, β_1 is constrained to be dependent upon predictor_2 by defining β_1 as a linear model itself and including predictor_2 in the definition. Using the previous multivariate regression equations, the linear interaction model takes the following form:

$$\begin{gathered} \text{outcome}_i \sim \text{Normal}(\mu_i, \sigma) \\ \mu_i = \alpha + \gamma_i \times \text{predictor}_{1,i} + \beta_2 \times \text{predictor}_{2,i} \\ \gamma_i = \beta_1 + \beta_{1,2} \times \text{predictor}_{2,i} \end{gathered} \tag{25}$$

Prior distributions are placed on all applicable parameters, but for simplicity have been omitted here. The parameter $\beta_{1,2}$ defines the strength of the dependency of outcome_i and predictor_1 on predictor_2; and γ_i is a placeholder that defines the linear function of the slope between

outcome_i and predictor_1. This equation defines the linear interaction effect between outcome_i and predictor_1. The interaction model creates posterior distributions that are conditional on those aspects of the data that posterior distributions from the simpler linear models cannot resolve. This modeling approach allows the relationship between the predictor variable and outcome to change depending upon another predictor variable, and provides the ability to estimate the aspects of a distribution of those changes.

Application to Gross Count Measurement Data

The scenario tested in this analysis is one in which outcome data either originated from measurements of ambient background or measurements with a radioactive source present. As outlined previously, the data analyzed were gross counts in fixed one second intervals, such that the variable $\text{outcome}_i = Count_i$. Suppose the experimenter wishes to study the relationship between a gross count measurement and the standard deviation of the gross counts obtained in the current measurement and the previous four measurements. This predictor can be designated as $SD5$, and $\text{predictor}_{1,i} = SD5_i$. The condition of whether the analyzed data are from a background measurement or a measurement with a source present requires the linear model to incorporate a categorical predictor, such that $\text{predictor}_{2,i} = Bkgd\_Sample_i$. In this categorical predictor, $Bkgd\_Sample = 1$ if $Count_i$ is attributed to a background measurement; 0 otherwise. The simplified multivariate linear model for this scenario takes the following form:

$$\begin{gathered} Count_i \sim \text{Normal}(\mu_i, \sigma) \\ \mu_i = \alpha + \beta_{SD5} \times SD5_i + \beta_{Bkgd\_Sample} \times Bkgd\_Sample_i \end{gathered} \tag{26}$$

Given this model and the strategy discussed previously regarding the components of an interaction model, we want $Count_i$ and $SD5_i$ to vary as a function of $Bkgd\_Sample_i$ This relationship is measured by the slope β_{SD5}. To accomplish the desired interaction, β_{SD5} is

constrained to being dependent upon $Bkgd\_Sample$ by defining a linear model for β_{SD5} that includes $Bkgd\_Sample$. The linear interaction model takes the following form:

$$\begin{gathered} Count_i \sim \text{Normal}(\mu_i, \sigma) \\ \mu_i = \alpha + \gamma_i \times SD5_i + \beta_{Bkgd\_Sample} \times Bkgd\_Sample_i \\ \gamma_i = \beta_{SD5} + \beta_{Bkgd\_Sample,SD5} \times Bkgd\_Sample_i \end{gathered} \tag{27}$$

By defining the relationship between $Count$ and $SD5$ with a linear interaction model, any change in μ_i resulting from a unit change in $SD5_i$ is given by γ_i. Now, to compute the relationship between $SD5_i$ and $Count_i$, incorporation of β_{SD5}, $\beta_{Bkgd\_Sample,SD5}$, and $Bkgd\_Sample_i$ is required. The model explicitly addresses the hypothesis that the slope between $Count$ and $SD5$ is conditional upon whether or not a measurement is from background, and the parameter $\beta_{Bkgd\_Sample,SD5}$ defines the strength of the dependency. Even though the parameters $SD5$ and γ are constructs of this experiment, in the Bayesian framework they can be thought of as targets of learning that can be characterized by a posterior density. This point highlights a valuable aspect of Bayesian statistical modeling: when an experimenter wishes to obtain abstract statistical knowledge about the data, a Bayesian procedure provides a model that can describe non-physical parameters. The presented linear interaction model allows for the visualization of how much the relationship between $Count$ and $SD5$ depends upon whether the measurement is background or not.

The full model including prior distributions takes the form:

$$\begin{gathered} Count_i \sim \text{Normal}(\mu_i, \sigma) \\ \mu_i = \alpha + \gamma_i \times SD5_i + \beta_{Bkgd\_Sample} \times Bkgd\_Sample_i \\ \gamma_i = \beta_{SD5} + \beta_{Bkgd\_Sample,SD5} \times Bkgd\_Sample_i \\ \alpha \sim \text{Normal}(\mu_\alpha, \sigma_\alpha) \\ \beta_{SD5} \sim \text{Normal}\left(\mu_{\beta_{SD5}}, \sigma_{\beta_{SD5}}\right) \\ \beta_{Bkgd\_Sample} \sim \text{Normal}\left(\mu_{\beta_{Bkgd\_Sample}}, \sigma_{\beta_{Bkgd\_Sample}}\right) \\ \beta_{Bkgd\_Sample,SD5} \sim \text{Normal}\left(\mu_{\beta_{Bkgd\_Sample,SD5}}, \sigma_{\beta_{Bkgd\_Sample,SD5}}\right) \\ \sigma \sim \text{Uniform}(\min_\sigma, \max_\sigma) \end{gathered} \tag{28}$$

This Bayesian model functions by using the built-in definitions in Equation 28 for the likelihood, the parameters, and the prior. These definitions are used to process the input measurement data, producing the posterior distribution. Essentially, the model is conditioning the prior on the data to approximate the posterior distribution. This conditioning is based on the rules of probability theory; namely Bayes' theorem and the product rule for probability theory (Gelman, et al., 2013). Conjugate models can be conditioned formally to produce a posterior distribution; however, the linear interaction model presented requires various techniques to approximate the mathematics that follow from the definition of Bayes' theorem. Given the use of normal distributions in the model, a quadratic approximation is sufficient to estimate the posterior distribution (Gelman, et al., 2013). Assuming that the posterior distribution $\mathrm{p}(\theta|y)$ is unimodal and nearly symmetric, it can be approximated by a normal distribution in that the logarithm of the posterior density is estimated by a quadratic function of θ. This technique takes advantage of the fact that the region near the peak of the posterior distribution will be nearly normal in shape in normal (Gaussian) mixture models. To improve the approximation, logarithms or logits of parameters are used to create a parabola that can be described by a quadratic function. An optimization algorithm in R locates the peak based on slopes between points, and the curvature near the peak is estimated. This curvature estimate is used to compute a quadratic approximation of the posterior distribution.

The model uses Bayesian updating, which allows the experimenter to consider every possible combination of values for the specified parameters (McElreath, 2016). Each of these combinations is scored by its relative plausibility, considering the data used. These relative plausibilities are the posterior probabilities for each of the combinations of parameter values. In other words, there are a finite number of normal distributions the model knows exist, and the

model considers every possible distribution defined by a combination of the input parameters. Consequently, the posterior distribution is a distribution of normal distributions constructed with the data and model in mind. The estimate provided by the model is then the posterior distribution itself, not a specific point within it.

Given the investigative nature of the experiment, appropriate selection of prior distributions and parameter values for the model is based on partial knowledge of the data to be analyzed and a preferred state of ignorance, premised on information theory and maximum entropy. The model used to generate the results in this paper takes the form:

$$\begin{gathered} Count_i \sim \text{Normal}(\mu_i, \sigma) \\ \mu_i = \alpha + \gamma_i \times SD5_i + \beta_{Bkgd\_Sample} \times Bkgd\_Sample_i \\ \gamma_i = \beta_{SD5} + \beta_{Bkgd\_Sample,SD5} \times Bkgd\_Sample_i \\ \alpha \sim \text{Normal}(600,10) \\ \beta_{SD5} \sim \text{Normal}(0,1) \\ \beta_{Bkgd\_Sample} \sim \text{Normal}(0,1) \\ \beta_{Bkgd\_Sample,SD5} \sim \text{Normal}(0,1). \\ \sigma \sim \text{Uniform}(0,50) \end{gathered} \tag{29}$$

Here, α is the expected value for $Count_i$ when all of the predictors are equal to 0. Based on knowledge from past observations, acceptable values for the expected gross counts, μ_i, from background measurements in the data analyzed in this investigation can be centered at 600 counts and assumed to be normally distributed over a relatively wide range. This allows the model to start at a reasonable approximation, as well as maintaining a partial state of ignorance. The values for the Gaussian priors of β_{SD5}, β_{Bkgd_Sample}, and $\beta_{Bkgd\_Sample,SD5}$ were selected because equal probability exists above and below 0, where a value of 0 suggests that the predictors have no relationship to the observed gross counts. This is a conservative assumption, especially considering that most of the probability mass is around 0. A flat prior for σ constructs a posterior distribution proportional to the likelihood, with the advantage that the data will drive the model towards an appropriate approximation.

The model operates by examining a set of previously recorded background gross count measurements, the training data Bkgd1, and resulting $SD5$ with a set of unknown sample gross count measurements and resulting $SD5$. These two sets of data make up the arrays for $Count_i$ and $SD5_i$. Included in the data array is $Bkgd\_Sample_i$, such that each index is correctly categorized as a known background measurement or an unknown sample measurement. This setup allows the model to work in a way that is intuitive to operational measurement technique: known background data and resulting estimates are used to create a relationship that is expected to be consistent across all measurements with no source present, and this relationship is compared to the samples in question to judge if a source is present. All statistical calculations and analyses were performed in the R statistical programming language and the Stan probabilistic programming language.

RESULTS AND DISCUSSION

Format of Investigation

An integral component of the results presented is the process required to validate a Bayesian model both statistically and operationally in Health Physics. The proceeding sections are formatted to demonstrate the method to develop a viable interaction model to detect radiological sources using gross counts with the parameter γ_{diff} with these principles in mind. The general structure of this procedure is:

- Interaction model validation
- Data exploration using Bayesian data analysis
- Production of an equivalent to the frequentist decision threshold using γ
- Application of the decision parameter to operational conditions and comparison to the frequentist decision threshold
- Testing and expansion of the model using established Bayesian statistical methods and tests
- Application of the results to operational considerations in Health Physics

This list by no means represents the only way to develop a decision threshold for gross counts using Bayesian statistics. The approach presented may, however, serve as a guideline for a series of requirements to integrate Bayesian modeling (specifically, an interaction model) with radiation detection using gross counts in low fidelity systems. By far the most crucial aspect of this process is contained within the last bullet, but the preceding sections in the list are necessary steps towards the requisite conclusions to be able to appropriately apply the model to an operationally relevant scenario in Health Physics. The first five sections of this chapter describe

the necessary model development and validation steps, while the last section describes the model application to the detection of radiological sources at low signal-to-background ratios.

Validating the Interaction Model

The initial step in the investigation to test the hypothesis that the relationship (slope) between $Count$ and $SD5$ is conditional upon whether or not the sample is from background is to analyze a large dataset with and without the interaction. Due to the partially non-informed prior distributions and parameter values in the model presented, it is important to assess the hypothesis in a manner that is driven completely by the data. This is achieved by using sufficient data that the non/partially-informed model has an insignificant impact on the resulting estimates other than shaping the results such that meaningful conclusions can be drawn. The clearest way to visualize the hypothesis is to analyze all 3600 (1800 from background and 1800 from unknown sample) measurement indices per comparison, plot the linear regression with and without the interaction model, and overlay the regression line of expected gross counts for a given $SD5$. These plots are shown in Figures 3-1, 3-2, 3-3, and 3-4. Figure 3-1 shows training background measurements (Background1/Bkgd1) versus 200 cm source distance measurements; Figure 3-2 shows training background measurements versus 400 cm source distance measurements; Figure 3-3 shows training background measurements versus the second set of background measurements (Background 2/Bkgd2); and Figure 3-4 shows the results of running the model against itself, where the training background measurements are also used as the unknown sample measurements. Plots on the left display linear regression using the training background data and plots on the right display linear regression using sample data. Clear differences are observed when comparing the top plots to the bottom plots within each figure. The top plots show a linear regression using the multivariate model without interaction, while the bottom plots show the

results from using the linear interaction model. The regression line is the expected gross counts for a given SD5, and the shaded region represents the credible interval.

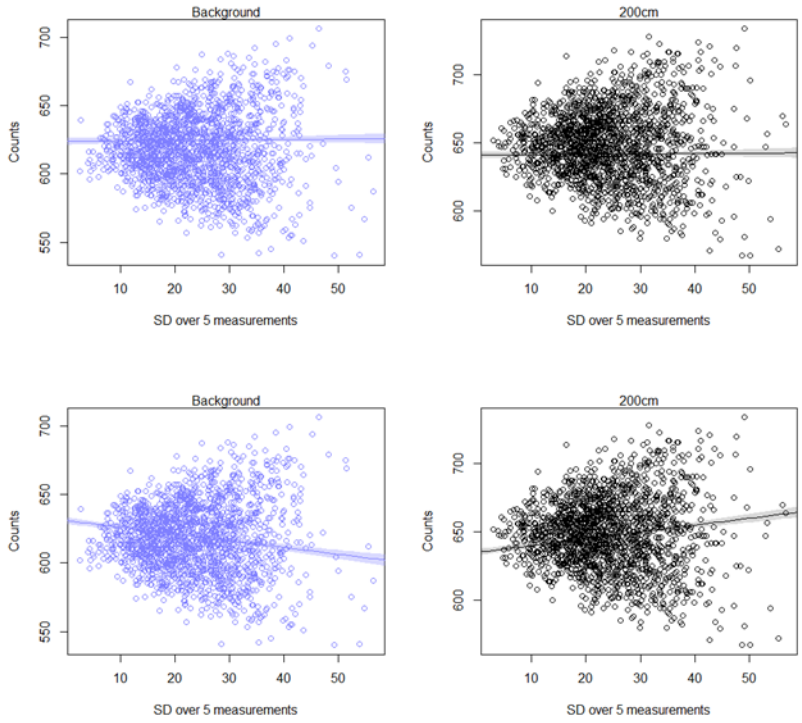

Figure 3-1. Regression plots for the 200cm source distance. The top plots show a linear regression using the multivariate model without interaction, while the bottom plots show the results from using the linear interaction model.

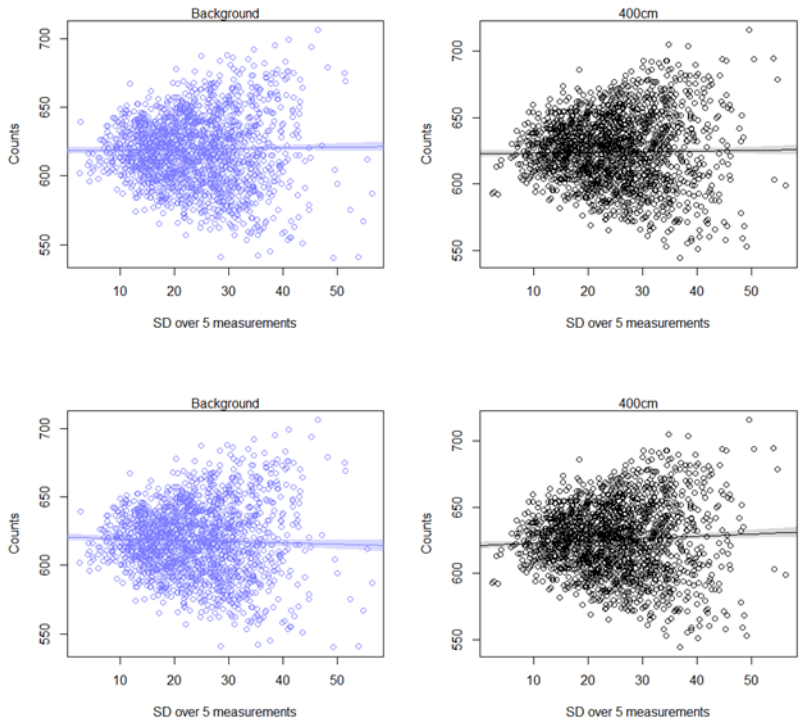

Figure 3-2. Regression plots for the 400cm source distance. The top plots show a linear regression using the multivariate model without interaction, while the bottom plots show the results from using the linear interaction model.

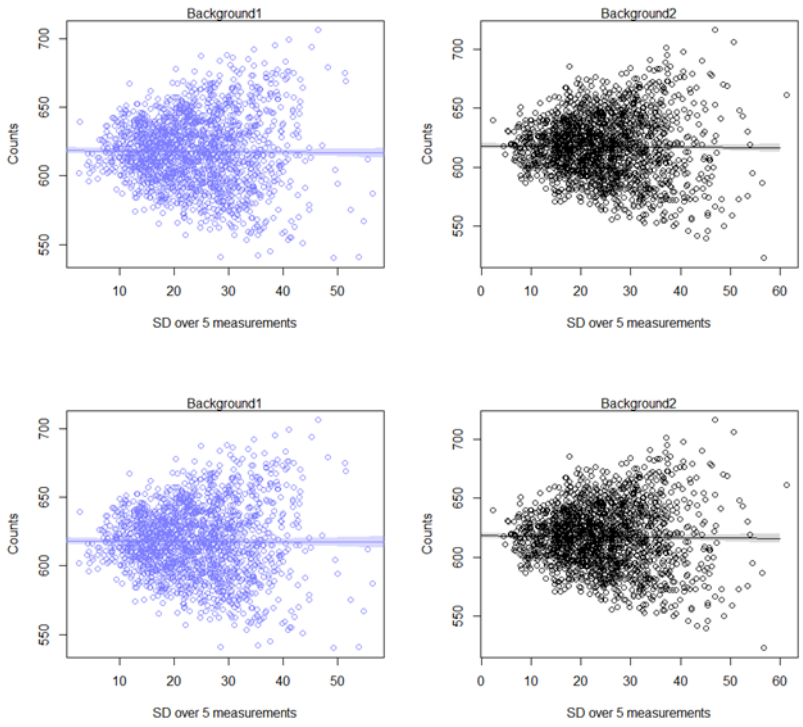

Figure 3-3. Regression plots for the Bkgd2 source distance. The top plots show a linear regression using the multivariate model without interaction, while the bottom plots show the results from using the linear interaction model.

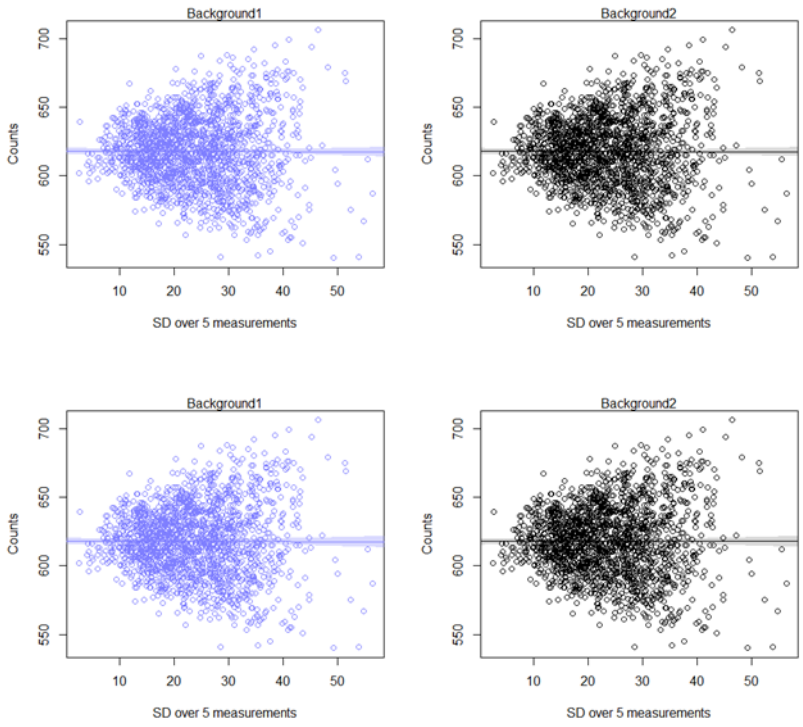

Figure 3-4. Regression plots for the Self-test. The top plots show a linear regression using the multivariate model without interaction, while the bottom plots show the results from using the linear interaction model.

In Figures 3-1 and 3-2, a clear difference exists between the slope of the line for the background and sample plots using the interaction. The positive slope seen in the 200 cm and 400 cm interaction plots suggests that a relationship between $Count$ and $SD5$ is conditional upon whether or not the sample is from background. To further substantiate this observation, it is apparent that when comparing Figures 3-3 and 3-4, the change in slope, if any, is minimal. This indicates that a relationship exists when comparing background measurements to sample measurements with a source present.

Symmetry in Interactions

Additional support for the interaction model observations is obtained by taking advantage of the fact that linear interactions are symmetrical (McElreath, 2016). Instead of suggesting that the relationship between $Count$ and $SD5$ depend upon whether or not the sample is from

background, we change the model such that we are investigating if the relationship between $Count$ and $Bkgd\_Sample$ is dependent upon $SD5$. To accomplish this, we still maintain every parameter and value from the original linear interaction model with the exception that we define a linear model for β_{Bkgd_Sample} that includes $SD5$. The model now takes the following form:

$$\begin{aligned} Count_i &\sim \text{Normal}(\mu_i, \sigma) \\ \mu_i &= \alpha + \gamma_i \times Bkgd\_Sample_i + \beta_{SD5} \times SD5_i \\ \gamma_i &= \beta_{Bkgd\_Sample} + \beta_{Bkgd\_Sample,SD5} \times SD5_i \end{aligned} \tag{30}$$

Essentially, we are asking if the gross counts obtained in background measurements versus measurements with a source present depend on the standard deviation of the gross counts obtained in the current measurement and the previous four measurements. Using the same data arrays and prior distributions from the linear regression experiment, the results for all four scenarios are provided in Figure 3-5.

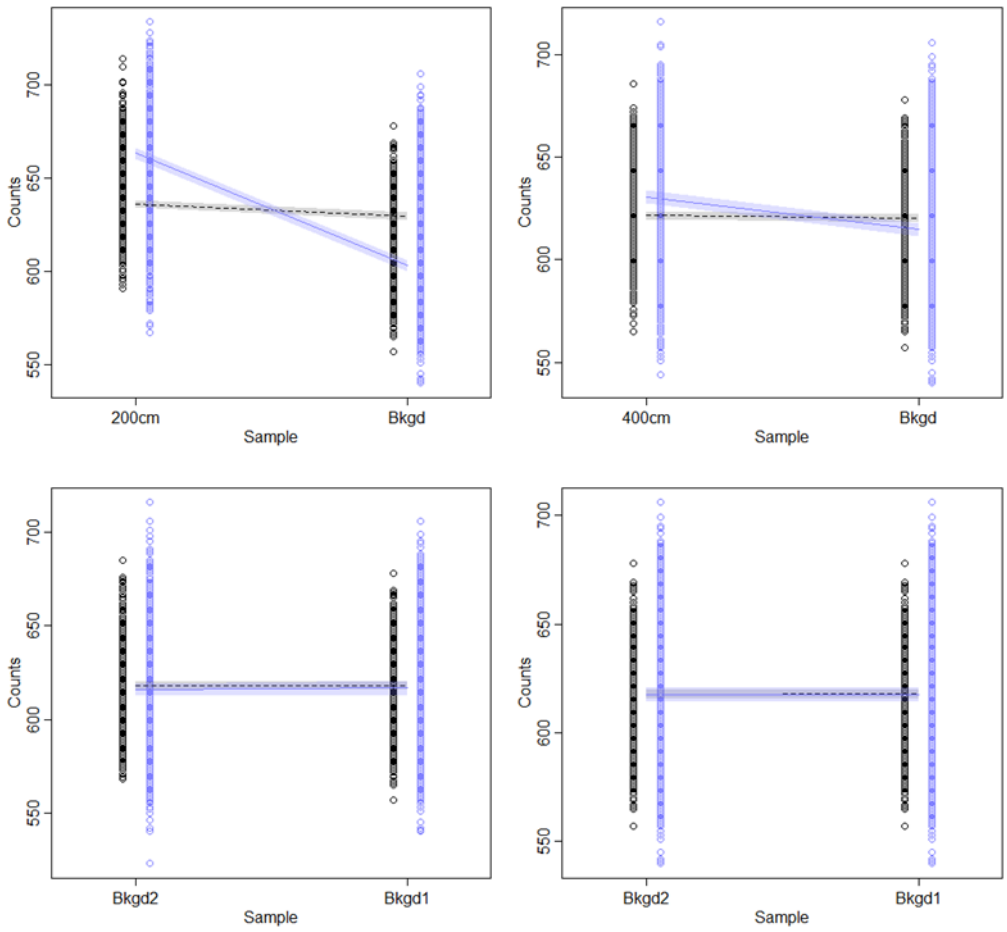

Figure 3-5. Symmetry plots comparing Bkgd1 (labeled Bkgd/Bkgd1) to the 200cm source distance (top left) dataset, the 400cm source distance dataset (top right), the Bkgd2 dataset (Bkgd2 in bottom left), and the Self-test (bottom right)

Within each plot, measurements from the training background data set are plotted on the right and the corresponding unknown sample measurements are plotted on the left. Within each category on the plots, the entries are split such that counts below the median $SD5$ value for that category are plotted on the left (black), and counts above the median $SD5$ value are plotted on the right (violet). The dashed line and shaded confidence region is the expected reduction in counts when we take an imagined sample with the minimum $SD5$ value and change between source present and no source present. It is apparent that across all scenarios, the slope of this line is nearly zero. This suggests that for any measurement, from background or otherwise, having a relatively low $SD5$ has a negligible effect on the gross counts for a particular measurement.

What is more important is the comparison of this line to the solid violet line within each plot. The solid line and shaded region represent the expected change in gross counts for an imaginary sample with a maximum $SD5$ value. This line shows a clear reduction in expected counts when compared between a measurement with a source present and no source present. As the comparison changes from a stronger (200 cm) to a weaker (400 cm) source to no source present (Bkgd2) versus the background measurement data (Bkgd1), this reduction decreases to a slope similar to the dashed line. These results indicate that for measurements associated with a high $SD5$, there is a positive effect (increase in expected gross counts) on the observed gross counts when a source is present. The original perspective presented, that the relationship between $SD5$ and gross counts in a measurement depends upon whether or not a source is present, is simultaneously true with this second perspective, that the relationship between whether or not a source is present and gross counts in a measurement is dependent upon $SD5$. This result establishes that the interaction model is valid. It is important to note however that these statements are only credible under the specific model assumptions and data used. To further confirm the linear interaction model, the same analyses were performed using two additional background measurement datasets (Bkgd2, Bkgd3) as the training dataset. The outcomes were similar to the previous results presented here. The results can be seen in Appendix A.

Using Bayesian Data Analysis to Explore Results

Given the confirmation of the model assumptions, the next step is to utilize Bayesian data analysis to approximate the posterior distributions of the relevant parameters to help understand the previous observations. Maximum a posteriori (MAP) estimates of the mean and standard deviation for the parameters α, β_{SD5}, β_{Bkgd_Sample}, $\beta_{Bkgd\_Sample,SD5}$, and σ were obtained using the quadratic approximation technique and are reported in Table 3-1.

Table 3-1. MAP Estimates for Parameters from the Interaction Model

	Unknown Sample Dataset							
	200cm		400cm		Background 2		Self	
parameter θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ
α	635.02	1.34	621.15	1.34	618.19	1.32	617.75	1.3
β_{SD5}	0.5	0.05	0.17	0.05	-0.04	0.05	0	0.05
β_{Bkgd_Sample}	-4.04	0.93	-0.75	0.93	-0.02	0.93	0.08	0.92
$\beta_{Bkgd\_Sample,SD5}$	-0.99	0.05	-0.27	0.05	0.02	0.05	0	0.05
σ	26.1	0.31	25.72	0.3	25.82	0.3	25.36	0.3

The table contains the estimates across all four comparison experiments: 200 cm source and 400 cm source distances, a separate background (Background 2/Bkgd2), andan identical dataset (Self). The parameter γ can be calculated using the relationship from Equation 27. For example, the MAP slope relating $SD5$ to gross counts for a measurement from background in the 200 cm experiment is

$$\gamma = \beta_{SD5} + \beta_{Bkgd\_Sample,SD5} \times 1 = 0.5 + (-0.99) = -0.49$$

and the MAP slope relating $SD5$ to gross counts for a measurement from a sample with the source present in the 200 cm experiment is

$$\gamma = \beta_{SD5} + \beta_{Bkgd\_Sample,SD5} \times 0 = 0.5 + 0 = 0.5$$

From this calculation, it is apparent that the relationship between $SD5$ and observed gross counts is reversed when comparing measurements from background and measurements with a source present. The calculated γ for each experiment is shown in Table 3-2.

Table 3-2. Calculated γ for Given Source Types

Source Type	γ	
	No Source	Source
200cm	-0.49	0.5
400cm	-0.1	0.17
Background 2	-0.02	-0.04
Self	0	0

The table shows that the relationship changes sign when the background dataset is compared to a dataset with a source present, whereas no reversal occurs when comparing the background dataset to another background dataset or itself. These results help explain the outcomes from the regression and symmetry figures and confirm that the relationship between $SD5$ and gross counts in a measurement depends upon whether or not a source is present. γ then provides a numerical value for that relationship.

Developing γ into a Decision Parameter

While all of the MAP estimate values provide verification that the model is producing sensible values based on the data, they do not provide any information regarding the uncertainty of the estimates. More specifically, to gain a better understanding of the relationship between $SD5$ and observed gross counts depending upon the presence of a source, we need to find the uncertainty associated with γ. This uncertainty can be obtained by computing the posterior distributions for the parameters in the model. These distributions can be approximated by sampling from the linear interaction model, which is achieved by sampling vectors of values from a multi-dimensional normal distribution. The resulting marginal distributions are shown in Figure 3-6; and the means, μ_γ, of these distributions are displayed in Table 3-3.

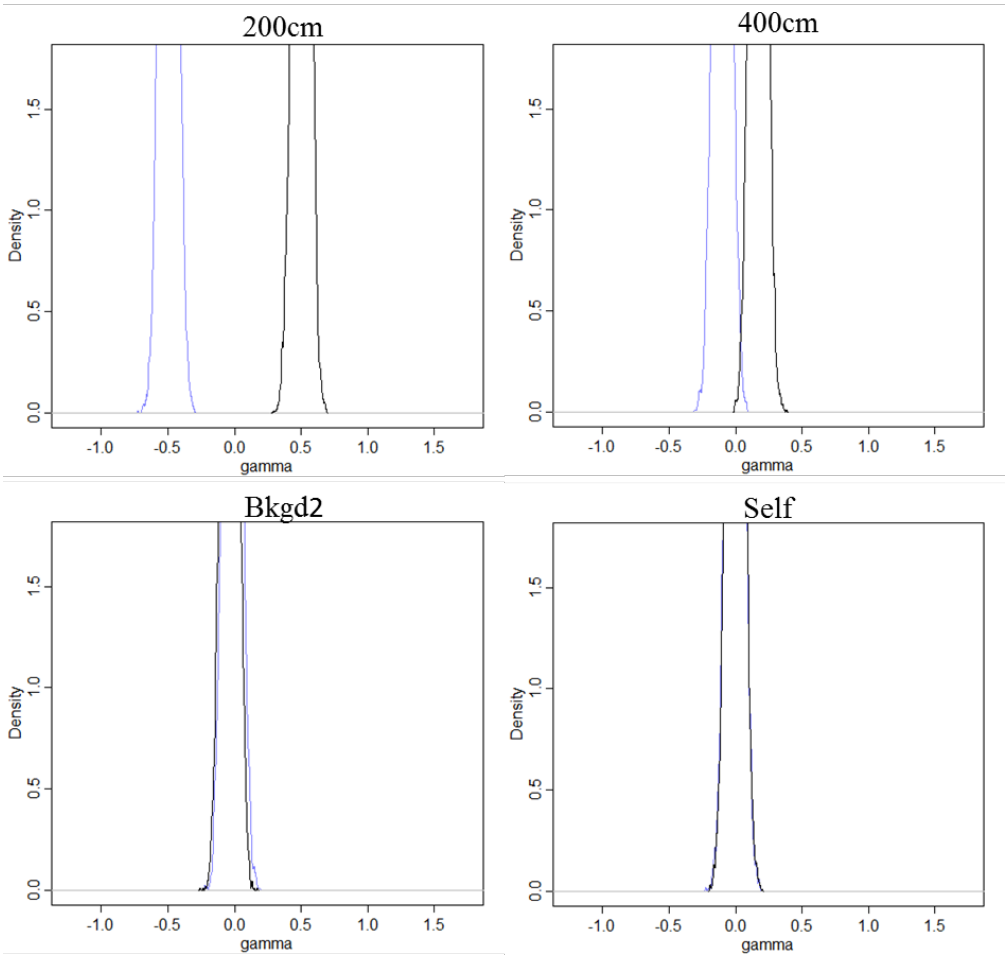

Figure 3-6. Comparisons of marginal distributions for Bkgd1 (violet) versus labeled source type (black)

Table 3-3. Means of the Marginal Distributions for a Given Source Type

Source Type	μ_γ	
	No Source	Source
200cm	-0.4943	0.5001
400cm	-0.1003	0.1715
Background 2	-0.0171	-0.0384
Self	-0.0053	-0.0021

The means are nearly equal to the MAP estimates for each corresponding calculated γ in Table 3-2, which is expected, and also confirms that the marginal distributions are valid. The tops of the marginal distributions are excluded from the figure due to the density being so relatively large that it would take away from the more important feature of the graphs: the distance

between the marginal distributions for the source type dataset (black) and the background training dataset (violet). Again, a clear trend is visible as a comparison is made from the stronger (200 cm) to the weaker (400 cm) source to no source present (Bkgd2) versus the background measurement data (Bkgd1).

The marginal distributions provide a response to the most useful question from the interaction model: what is the probability that the relationship between $SD5$ and gross count measurements from background is less than the relationship between $SD5$ and gross count measurements from a sample? This question can be answered by computing the difference between the marginal distributions, γ_{diff}, for each sample from the posterior, and find the proportion of differences below 0. A total of 10^4 samples are taken from the posterior for each γ_{diff} calculated. The distributions for each γ_{diff} can be seen in Appendix A. The results of these computations for each of the source strengths are shown in Table 3-4.

Table 3-4. γ_{diff} for Given Source Types

Source Type	γ_{diff}
200cm	1
400cm	1
Background 2	0.3201
Self	0.5284

These values are not the same as the overlap of the marginal distributions in Figure 3-6, because the values are derived from the posterior distribution of differences in slopes. It is important to remember that this Bayesian model structure produces distributions of distributions, which is why this distinction exists and what makes γ_{diff} unique. Additionally, these are not the probability of any observable event. They are probabilities computed by the model based on the question asked, given this particular dataset. For example, of all of the possible states the model calculates in the experiment with the 200 cm source distance dataset, 100% are consistent with

the data and claim that γ_{Bkgd} is less than γ_{200cm}. In repeat sampling, we do expect 100% of the measurements from background to have a γ less than measurements with a source present at a distance of 200 cm from the detector. This is an important concept to distinguish when drawing conclusions from a Bayesian model. The model serves as a way to update the operators' beliefs given their previous assumptions and the data presented; it does not directly answer the question "what is the probability of observing an event?" This is what makes comparing observable outcomes from a Bayesian analysis and a frequentist approach problematic. However, decisions based on those observable outcomes can be compared, especially in a simplified scenario where the decision space is either detection or no detection. Using the linear interaction model as a detection method can be achieved by setting a threshold on the probability of γ_{diff}. Doing so creates an operationally acceptable limit on how likely it is that the relationship between observed gross counts and $SD5$ from measurements of an unknown sample is greater than the relationship between observed gross counts and $SD5$ from measurements of a background sample. Detection using γ_{diff} is then accomplished by defining a decision such that a set of measurements, i, analyzed in the model are considered to trigger an alarm if

$$\Pr^* < \gamma_{diff_i} \tag{31}$$

where $\Pr^*$ is a pre-specified, fixed probability and γ_{diff_i} is the probability that the relationship between observed gross counts and $SD5$ from measurements of an unknown sample are greater than the relationship between observed gross counts and $SD5$ from measurements of a background sample.

Testing Under Operationally Equivalent Conditions

The results presented up to this point have been based on analyses using 1800 one s gross count measurements. The total time required to collect this data is 30 minutes, which is a relatively long detection time. Considering that $SD5$ is constructed from 5 individual 1 s gross count measurements, it is reasonable to test the model using a data array of $i = 5$. This would be equivalent to a 5 s measurement time consisting of $n = 5$ individual measurements. Before this can be done, the model's output must be tested at reducing n values to verify that the observations seen in the $n = 1800$ experiment would hold. The regression plots, symmetry plots, MAP estimates, calculated γ tables, marginal distributions, and marginal μ_γ tables for these n values can be seen in Appendix B. γ_{diff} for each n value per source type is shown in Table 3-5.

Table 3-5. γ_{diff} per n Measurements for a Given Source Type

n	1800	900	450	225	113	57	29	15	8	4
200cm	1	1	1	1	1	1	0.999	0.966	0.985	0.699
400cm	1	1	1	1	0.982	0.74	0.791	0.789	0.614	0.216
Bkgd2	0.32	0.486	0.566	0.605	0.302	0.201	0.555	0.079	0.065	0.052
Self	0.528	0.51	0.517	0.509	0.496	0.478	0.47	0.437	0.433	0.42

It should be noted that the data used per n overlap, meaning that the data from $n = 4$ make up half of the data from $n = 8$, and the data from $n = 8$ make up half of the data from $n = 15$, etc. This was done to see how the relationships modeled from the data changed as n was reduced. If non-overlapping data were used, it would be difficult to discern if the outcomes observed were a result of n decreasing, or if the outcomes were influenced by the unique relationship between observed gross counts and $SD5$ for that particular dataset. It can be seen that similar trends from the $n = 1800$ dataset persist as n is reduced, confirming that testing the model with an $n = 5$ (or equivalently, $i = 5$) dataset is acceptable. The Self results are always expected to be somewhere

around $\gamma_{diff} = 0.5$ due to the model not being able to find a significant difference between the two datasets being analyzed, evident by the fact that both categories ($Bkgd\_Sample_i = 0, 1$) are being given roughly equal probability densities by the model.

Commonly in non-Bayesian statistical inference, a minimum number of observations are required for a useful statistical estimate, making statistical inferences questionable at small sample sizes. A Bayesian estimate, in contrast, is valid at any sample size. This does not mean that more observations are not useful; the estimates just have a clear interpretation based on the model. In the case of only using five measurements per analysis in the model, the same question is asked before using γ_{diff}, but basing the question on considerably fewer samples. The model is approximating the relationship between $SD5$ and gross counts based on whether or not a source is present using only the five measurements analyzed and the inferences used to shape the model (the priors). The strength of γ_{diff} is that the uncertainty associated with using only five measurements is contained entirely in the marginal distributions for γ_{Bkgd} and $\gamma_{Unknown}$. The only uncertainty that can be associated with γ_{diff} is sampling uncertainty, which is addressed by using a large number of samples (10^4) taken from the posterior distribution to approximate the marginal distributions used to calculate γ_{diff}.

The same data used to generate the 1800 individual 1 s measurement results were used, but the data were partitioned into five 1 s measurement sequences. Figure 3-7 displays the results from analyzing 360 independent measurement samples using Pr^* values ranging from 0.5 to 0.9975.

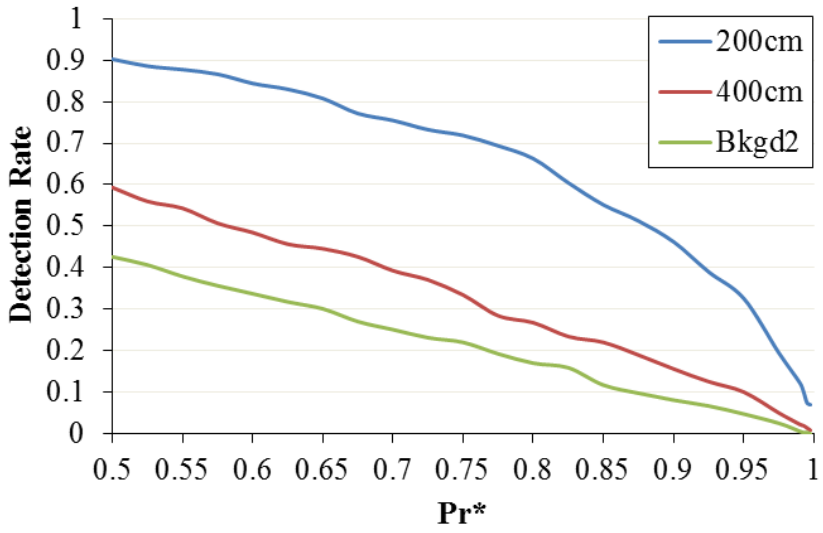

Figure 3-7. Detection rates for different Pr*

While the gross count measurements used to calculate an $SD5$ in a given measurement sequence are being used in the preceding measurement sequence as outcome variables in the regression, this overlap is equivalent to an online measurement system. Detections for the 200 cm and 400 cm source distances can be equated to a true detection rate (true positives), and detections on Bkgd2 are equivalent to false alarms (false positives). Setting $\mathrm{Pr}^{*} = 0.5$ is the least conservative decision, evident from the amount of alarms triggered on the Bkgd2 data. At an approximate "false positive" rate of 0.05 (when the detection rate of Bkgd2 ≈ 0.05), the detection rate for the 200 cm source distance is $\sim$0.33, and $\sim$0.10 for the 400 cm source distance. This is roughly equal to the alarm rates observed using the frequentist decision threshold on Page 20. To further relate these results to common frequentist data analyses, characteristic limits for measurements of ionizing radiation (Equation 4) were applied to the same data. The data were transformed such that they would be in the form of gross count measurements over 5 s intervals, equivalent to the amount of time used for one "measurement" dataset analyzed by the interaction model. The decision threshold was determined using the same background measurement data utilized as the training dataset in the Bayesian interaction

model. Notably, the entire dataset was used to determine this decision threshold. The interaction model only analyzes the equivalent of one 5 s background measurement each time it performs its analysis, under a partially non-informed prior. However, basing decisions on one 5 s background measurement is not good measurement technique using frequentist statistics, and all 360 measurements were used to determine the decision threshold. This decision threshold was used to analyze data from the 200 cm and 400 cm source distances to determine the fraction of true positive detections on the data. The same decision threshold was used to determine the fraction of alarms on the Bkgd2 data. This was done so that results from the decision threshold analysis could be directly compared to results from the interaction model analysis. Figure 3-8 shows the comparison of true positives to false positives for the interaction model and the decision threshold analyses at operationally relevant false positive rates (false positives ≤ 0.1).

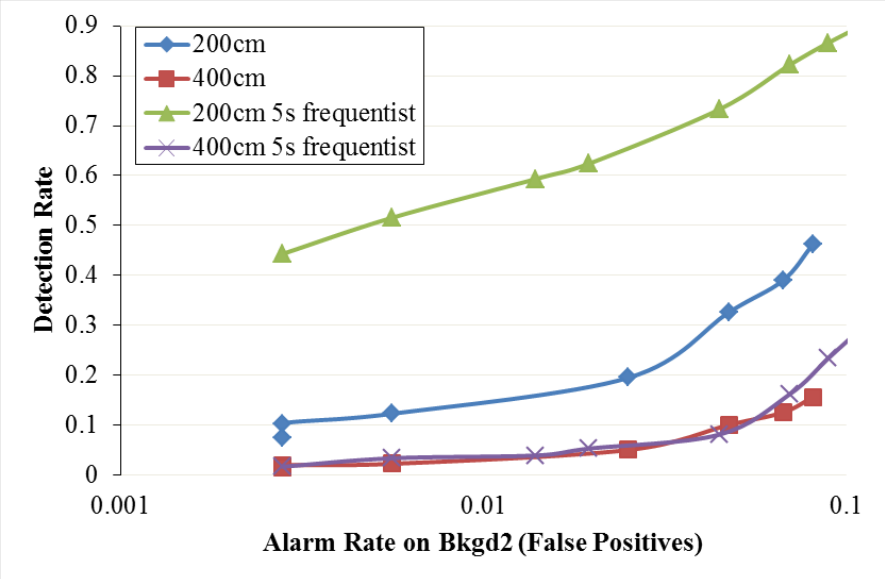

Figure 3-8. Detection efficiencies for the interaction model and the frequentist decision threshold (5s traditional) for the 200cm and 400cm source distances for false positives < 0.1 (log scale)

Clearly, the frequentist decision threshold outperforms the interaction model when comparing the 200 cm source distance data, but they are similar when comparing the 400 cm source distance data. This result is encouraging, considering that detection of weaker sources is

challenging for radiation counting statistics. Further, the frequentist decision threshold was not able to attain a false positive rate equal to 0, while the interaction model was able to do so and still had positive detections on the sources (0.070 for the 200 cm source distance and 0.008 for the 400 cm source distance). The limitations of the dataset size may have influence on results for lower false positive rates. Given that the dataset consists of 360 observations, having one false positive is equal to a false positive rate $= 1/360 = 0.0028$. This false positive rate occurred for two different Pr^* values (0.995 and 0.9925), whereas this only occurred for $\alpha = 0.0025$ for the frequentist decision threshold. More data could be collected to further study the difference between the frequentist decision threshold and Bayesian interaction model. However, considering no significant difference between the two approaches is observed with the current 400 cm dataset, no further data inclusion is deemed necessary at this time as significant differences are not expected.

Model Specification and Optimization

An essential facet of Bayesian data analysis is model specification and optimization (Gelman, et al., 2013). Bayesian data analysis requires the construction and testing of models that are applied to the data in question. This is due to the fact that infinite models can be created to describe a dataset; however, only a small fraction of those are applicable to the research question in a logical manner. The idea is to build models with improved parameter estimates to better describe the phenomena being analyzed and test their estimated predictive efficiencies with a WAIC test. Applied to the work presented thus far, the simplest way to do this is test the original multivariate model with no interaction against interaction models with identical distribution forms but different parameter bounds. This also satisfies a basic tenant in Bayesian model development: *models with fewer assumptions are to be preferred* (Michel, 2016). The

original interaction model contained normal distributions with parameter estimates that maintained a state of partially-informed ignorance. Changing the distributions used in the original model would add significant assumptions to the model, as well as increase the complexity of model selection. Two viable assumptions that could be applied to this model are using a model with an informed α-intercept, and using a model that allows broader values for linear relationships. An informed α-intercept model would force the model to focus on a specific range for the expected gross count values when data inputs are only from background. Ideally, the α prior distribution would possess a mean value equal to the mean of the background gross count distribution training dataset (617.7) with a relatively narrow standard deviation. Adhering to these concepts, the model takes the following form:

$$\begin{gathered} Count_i \sim \text{Normal}(\mu_i, \sigma) \\ \mu_i = \alpha + \gamma_i \times SD5_i + \beta_{Bkgd\_Sample} \times Bkgd\_Sample_i \\ \gamma_i = \beta_{SD5} + \beta_{Bkgd\_Sample,SD5} \times Bkgd\_Sample_i \\ \alpha \sim \text{Normal}(617.7,3) \\ \beta_{SD5} \sim \text{Normal}(0,1) \\ \beta_{Bkgd\_Sample} \sim \text{Normal}(0,1) \\ \beta_{Bkgd\_Sample,SD5} \sim \text{Normal}(0,1). \\ \sigma \sim \text{Uniform}(0,50) \end{gathered} \tag{32}$$

The only difference between this model and the original interaction model in Equation 29 is the parameter values for the α prior. A standard deviation of 3 was chosen so that the range of values were limited, but not so stringent that the model would undoubtedly cause the data to be fit incorrectly. The second variation of the original model, where broader values for linear relationships are applied, is constructed by increasing the standard deviation for all β priors. The model takes the following form:

$$
\begin{gathered}
Count_i \sim \text{Normal}(\mu_i, \sigma) \\
\mu_i = \alpha + \gamma_i \times SD5_i + \beta_{Bkgd\_Sample} \times Bkgd\_Sample_i \\
\gamma_i = \beta_{SD5} + \beta_{Bkgd\_Sample,SD5} \times Bkgd\_Sample_i \\
\alpha \sim \text{Normal}(600,10) \\
\beta_{SD5} \sim \text{Normal}(0,2) \\
\beta_{Bkgd\_Sample} \sim \text{Normal}(0,2) \\
\beta_{Bkgd\_Sample,SD5} \sim \text{Normal}(0,2). \\
\sigma \sim \text{Uniform}(0,50)
\end{gathered} \tag{33}
$$

A value of 2 was chosen because, in the testing process, if this model was preferred over the original model then the experimenter would know that the linear relationship is indeed larger than initially believed.

Infinite ranges of values could be applied in different combinations to these two basic variations of the original model, but the point of model validation tests is to help steer the experimenter in a logical direction. In theory, operators could test as many models for estimated predictive efficiency as they desire, but drawing meaningful conclusions from the WAIC test results would be difficult. Choosing these models (multivariate linear regression with no interaction from Equation 26, original interaction model, informed α-intercept, and larger β relationship) allows the operators to test whether or not the interaction model is preferred; and if it was preferred, should they be more stringent in telling the model what to anticipate for the expected gross count value, or is the relationship between outcomes and predictors larger than believed. Lastly, because the desired application of this model is to be able to test datasets in operationally relevant scenarios, the operators want to determine what models have a higher estimated predictive efficiency at reduced n observations. WAIC testing was performed on these four models for $n = 1800$, $n = 900$, $n = 225$, $n = 29$, and $n = 5$ observations. For nomenclature simplicity, m7.4 is the multivariate linear regression model with no interaction, m7.5 is the original interaction model, m7.5a is the informed α-intercept model, and m7.5b is the larger β relationship model. The results tables below report the Akaike weight, which is "an

estimate of the probability that the model will make the best predictions on new data, conditional on the set of models considered" (McElreath, 2016). The training background dataset Bkgd1 was tested with the 200 cm dataset, the 400 cm dataset, and the Bkgd2 dataset.

Table 3-6. 200cm Dataset Akaike Weights from WAIC Test

n	1800	900	225	29	5
m7.5	0	0	0.3	0.27	0.29
m7.5a	0	0	0	0.49	0.01
m7.5b	1	1	0.7	0.24	0
m7.4	0	0	0	0	0.7

Table 3-7. 400cm Dataset Akaike Weights from WAIC Test

n	1800	900	225	29	5
m7.5	0.32	0.33	0.29	0.11	0.27
m7.5a	0.24	0.36	0.45	0.35	0.45
m7.5b	0.44	0.3	0.26	0.1	0.02
m7.4	0	0	0	0.44	0.26

Table 3-8. Bkgd2 Dataset Akaike Weights from WAIC Test

n	1800	900	225	29	5
m7.5	0.18	0.17	0.15	0.2	0.37
m7.5a	0.22	0.23	0.25	0.06	0.02
m7.5b	0.17	0.15	0.15	0.21	0.07
m7.4	0.44	0.45	0.44	0.53	0.55

In Tables 3-6 and 3-7, tests where a source was present in the dataset, no model is preferred across all n observation sizes; however, m7.4 only possesses weight in 3 out of the 10 tests run. In Table 3-8, m7.4 is preferred across all n observation sizes. This result is expected because the test indicates that a model without interaction is preferred when two datasets from background are tested in the model. In this instance, no interaction model would be required. These results support the results prior to the WAIC testing claiming that the relationship between $SD5$ and gross counts in a measurement depends upon whether or not a source is present. Even though no

model was preferred outright in the WAIC test when a source was present, it does not mean that an interaction model is not a valid model to apply to this scenario and dataset. It points out that none of these models conclusively outperform each other on estimated future data prediction given the assumptions. There are two instances where m7.4 is preferred when a source is present, but this result only occurs at $n = 29$ and $n = 5$. For larger n values, in which sufficient data are present to accurately describe the relationship between $SD5$ and gross counts in a measurement, interaction models are preferred. Given this result, the selection of m7.4 at these smaller n observation sizes suggests that the interaction models tested do not describe the relationship between $SD5$ and gross counts in a measurement for that particular dataset accurately enough. To put this result into perspective, the γ_{diff} values for the 200 cm, 400 cm, and Bkgd2 datasets for the $n = 5$ observations are 0.857, 0.327, and 0.025, respectively. In the 200 cm WAIC test for $n = 5$, m7.4 was preferred even though the γ_{diff}probability was relatively high. This suggests that WAIC testing may not be a viable metric for determining detection efficiency. Another factor to consider is that the datasets with a size of $n = 5$ will vary widely and produce a range of values for γ_{diff}, therefore each model must be tested to ascertain their detection efficiencies.

To determine the detection efficiency of the models, each of the models was tested with the same dataset used to test the original interaction model. The non-interaction model was excluded because it does not produce a posterior distribution for γ, and no detection decision can be made. A model that combined the informed α-intercept and larger β relationship (termed m7.5ab) was also tested to see if the two assumptions functioned "synergistically" and increased the detection rate when compared to other models. This model has the following form:

$$
\begin{gathered}
Count_i \sim \text{Normal}(\mu_i, \sigma) \\
\mu_i = \alpha + \gamma_i \times SD5_i + \beta_{Bkgd\_Sample} \times Bkgd\_Sample_i \\
\gamma_i = \beta_{SD5} + \beta_{Bkgd\_Sample,SD5} \times Bkgd\_Sample_i \\
\alpha \sim \text{Normal}(617.7,3) \\
\beta_{SD5} \sim \text{Normal}(0,2) \\
\beta_{Bkgd\_Sample} \sim \text{Normal}(0,2) \\
\beta_{Bkgd\_Sample,SD5} \sim \text{Normal}(0,2). \\
\sigma \sim \text{Uniform}(0,50)
\end{gathered} \tag{34}
$$

Total detections on the 200 cm, 400 cm, and Bkgd2 datasets for given Pr* values are shown in Appendix C. Figures 3-9 and 3-10 display results in a manner identical to Figure 3-8, where the fraction of detections with a source present (Detection Rate) was plotted against the fraction of detections on Bkgd2 (False Positives). The frequentist decision threshold (5s frequentist) detection rates are included to make the results operationally relevant.

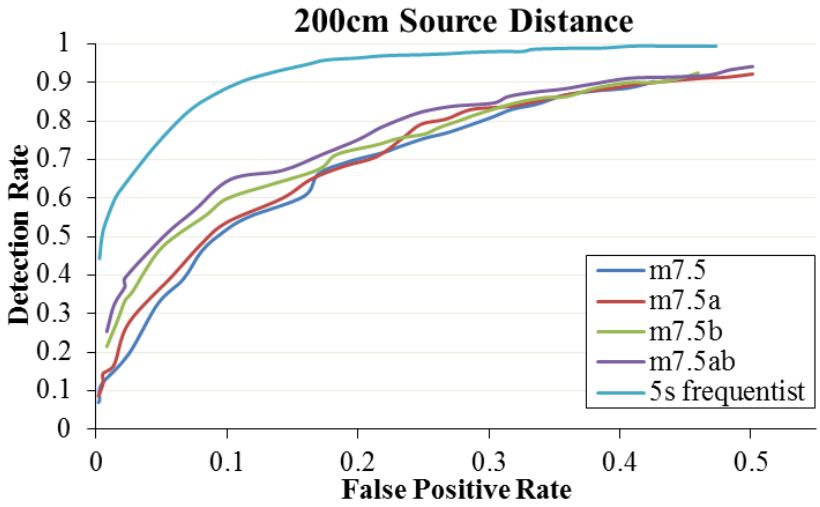

Figure 3-9. Detection efficiencies for the various interaction models and the frequentist decision threshold (5s traditional) for the 200cm source distance

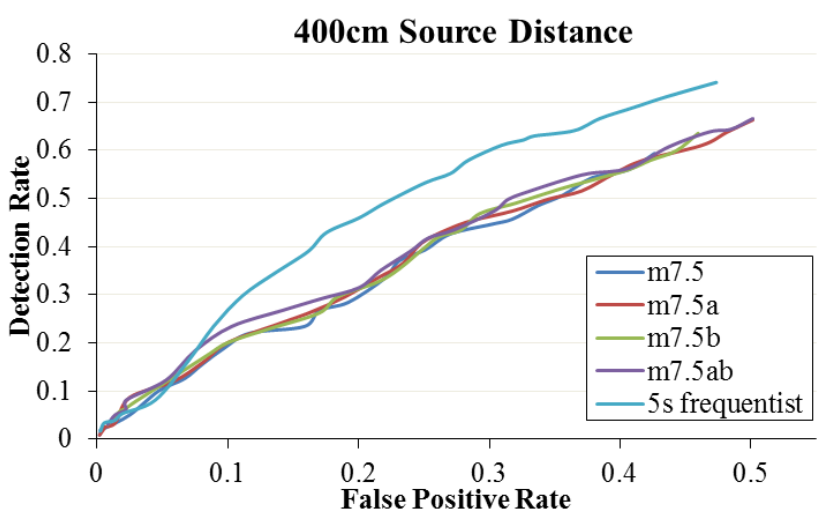

Figure 3-10. Detection efficiencies for the various interaction models and the frequentist decision threshold (5s traditional) for the 400cm source distance

The frequentist decision threshold clearly outperforms the interaction models for the 200 cm source distance dataset. The original interaction model (m7.5) performs the worst when compared to the other interaction models. The same trends are not evident in the 400 cm source distance dataset. While the frequentist approach does outperform the interaction models at larger false positive rates, the interaction models appear to perform better at lower false positive rates (false positives < 0.1). Figure 3-11 shows the plot for detections at lower false positive rates (log scale).

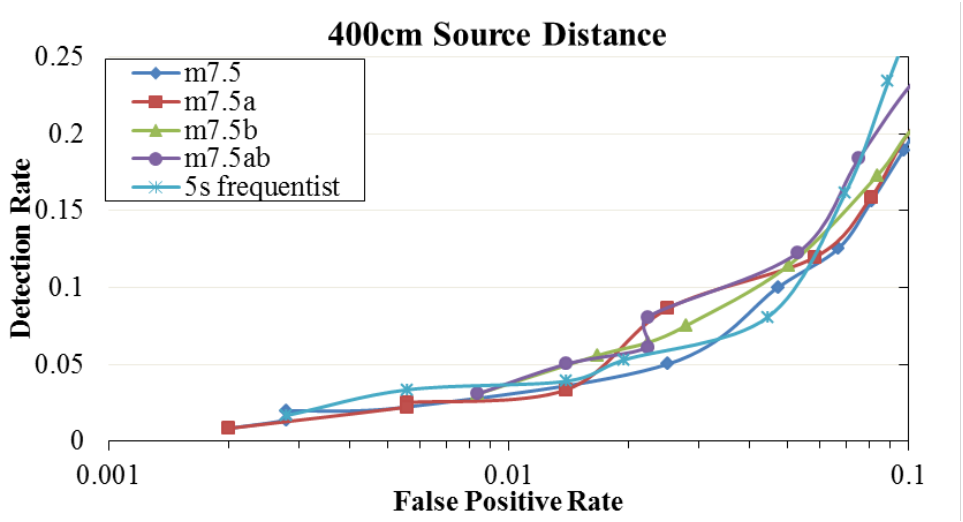

Figure 3-11. Detection efficiencies for the various interaction models and the frequentist decision threshold (5s traditional) for the 400cm source distance for false positives < 0.1 (log scale)

Operationally, these are the most important values to consider because it is rare to use a detection method allowing greater than 10% false positives. Although no model conclusively performs the best, it can be seen that there are instances where the interaction model is performing better for a given alarm rate on Bkgd2. Again, these results may be limited at lower false positive rates by the $n = 360$ observations used.

Another possible influence on the detection efficiency may be the manner in which the categorical predictor is assigned. In all of the models presented thus far, the categorical predictor, $Bkgd\_Sample = 1$ if $Count_i$ is attributed to a background measurement; 0 otherwise. Models were developed such that the categorical predictor was flipped in which 0 is assigned to background measurements. The same model forms (m7.5, m7.5a, m7.5b, m7.5ab) as before were tested with this flipped predictor and termed m7.5_0, m7.5a_0, m7.5b_0, and m7.5ab_0. Figures 3-12, 3-13, and 3-14 show the results.

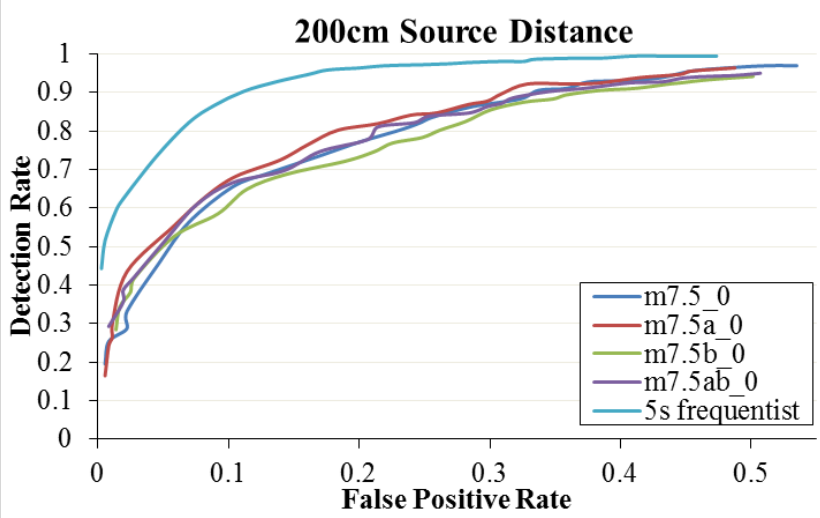

Figure 3-12. Detection efficiencies for the various flipped categorical predictor interaction models and the frequentist decision threshold (5s traditional) for the 200cm source distance

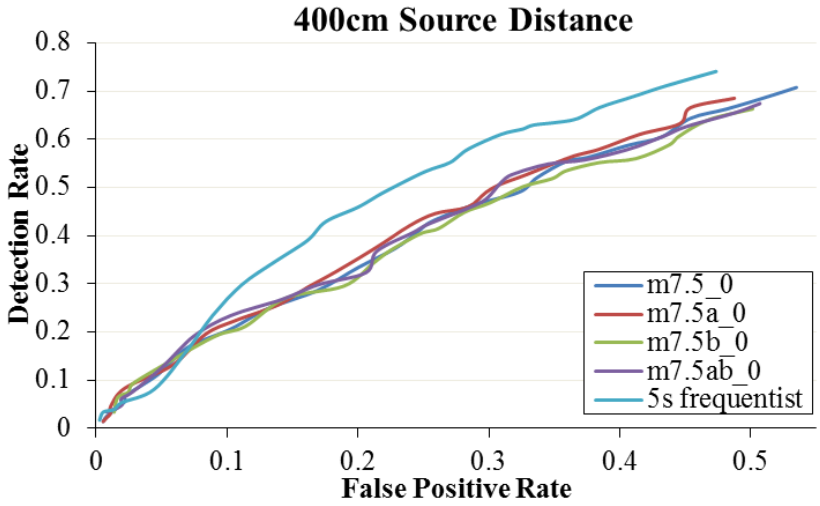

Figure 3-13. Detection efficiencies for the various flipped categorical predictor interaction models and the frequentist decision threshold (5s traditional) for the 400cm source distance

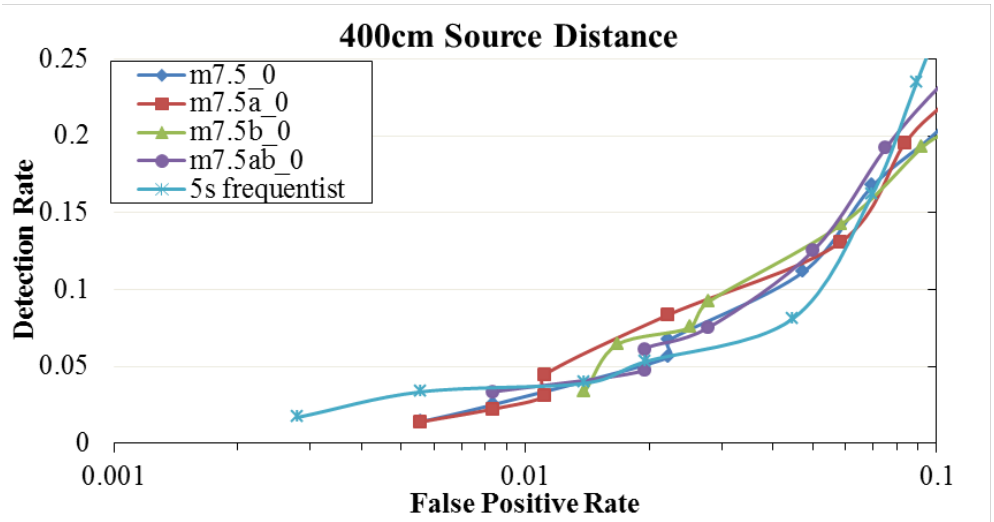

Figure 3-14. Detection efficiencies for the various flipped categorical predictor interaction models and the frequentist decision threshold (5s traditional) for the 400cm source distance for false positives < 0.1 (log scale)

The same datasets were used (except for the change in the categorical predictor in the data array) as in the previous results in Figures 3-9 through 3-11. Similar trends are observed between the two experiments, with the exception that the fraction of detections has increased for the flipped categorical predictor models at higher false positive rates (Alarms on Bkgd2 > 0.1). However, at operationally relevant false positive rates, the results are similar to those shown in Figure 3-11. It can be seen that flipping the predictor does not have a significant impact on detection efficiency for this dataset, and the model results are not influenced by the designation of the categorical predictor variable. Total detections on the 200 cm, 400 cm, and Bkgd2 datasets for given $\Pr^*$ values can be seen in Appendix C.

Embedded in all of the models presented up to this point is the assumption that the relationship between $SD5$ and gross counts in a measurement is linear. A power model was created and tested to see if the detection efficiency increased under this assumption. Equation 27 was modified so that the relationship between $SD5$ and gross counts in a measurement was taken to the 1.5 power, such that:

$$Count_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \gamma_i \times SD_i^{1.5} + \beta_{Bkgd\_Sample} \times Bkgd\_Sample_i$$
$$\gamma_i = \beta_{SD5} + \beta_{Bkgd\_Sample,SD5} \times Bkgd\_Sample_i \quad (35)$$

The model was tested with the exact same parameters as m7.5, m7.5a, m7.5b, and m7.5ab with the exact same dataset. The results are also plotted in an identical manner in Figures 3-15, 3-16, and 3-17. The same model names used in the linear model are reported in these figures to make comparisons simpler.

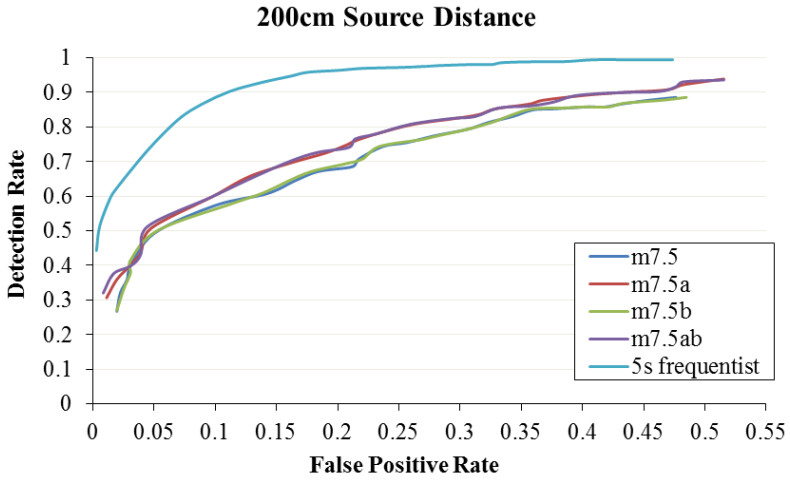

Figure 3-15. Detection efficiencies for the various power function interaction models and the frequentist decision threshold (5s traditional) for the 200cm source distance

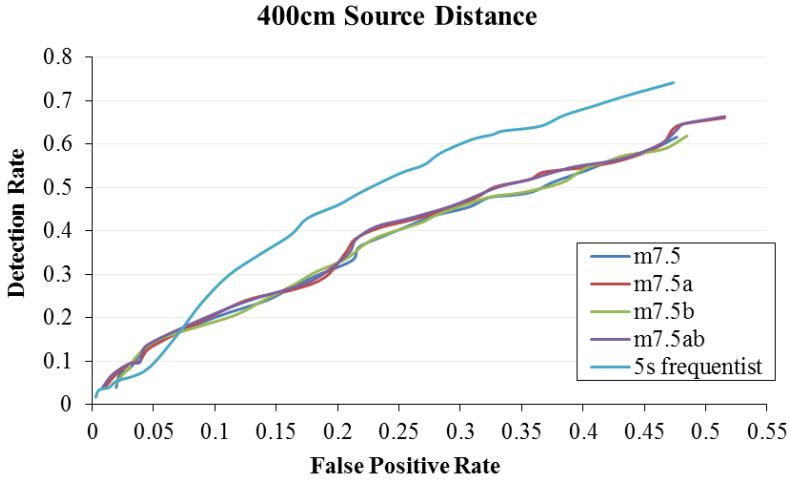

Figure 3-16. Detection efficiencies for the various power function interaction models and the frequentist decision threshold (5s traditional) for the 400cm source distance

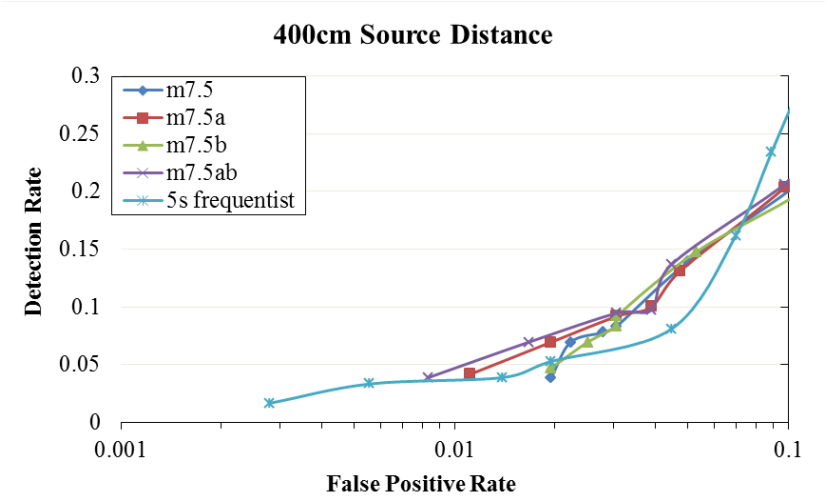

Figure 3-17. Detection efficiencies for the various power function interaction models and the frequentist decision threshold (5s traditional) for the 400cm source distance for false positives < 0.1 (log scale)

As with the other models, the 200 cm dataset heavily favors the frequentist decision threshold. The 400 cm dataset results are comparable to the linear model results, suggesting that changing the model from a linear to a power model does not significantly change the detection efficiency of the interaction model.

Simple model alterations, like the power model presented above, did not change the calculation of γ_{diff} considerably because the assumptions for the distributions and parameter estimates did not change. This allows the continued use of the quadratic approximation to calculate the necessary estimates for the relevant posterior distributions. A noted strength in Bayesian statistics is the development of statistically diverse models that could not be applied in the frequentist perspective. The models typically take on forms involving non-conjugate priors and hyperparameters (Gelman, et al., 2013) that require MCMC to estimate the posterior distribution. Such multilevel models require diagnostic tools to ensure that Markov chains have been defined correctly and converge to an estimate, that each successive sample within each parameter is not highly correlated with the sample before it, and that the chains are efficient

samplers. While these are valuable tools for a Bayesian statistician, MCMC algorithm performance is outside of the scope of this research. Further, the philosophical implications for creating a diverse model applied to this detection scenario require statistical theory far beyond what has been presented in this dissertation. However, a simple multilevel model can be constructed that maintains the entropy and information criterion principles. This is accomplished by applying hyperparameters to the α-intercept. The model takes the following form

$$
\begin{gathered}
Count_i \sim \text{Normal}(\mu_i, \sigma) \\
\mu_i = \alpha + \gamma_i \times SD5_i + \beta_{Bkgd\_Sample} \times Bkgd\_Sample_i \\
\gamma_i = \beta_{SD5} + \beta_{Bkgd\_Sample,SD5} \times Bkgd\_Sample_i \\
\alpha \sim \text{Normal}(\mu_\alpha, \sigma_\alpha) \\
\mu_\alpha \sim \text{Normal}(600,10) \\
\sigma_\alpha \sim \text{HalfCauchy}(0,1) \\
\beta_{SD5} \sim \text{Normal}(0,1) \\
\beta_{Bkgd\_Sample} \sim \text{Normal}(0,1) \\
\beta_{Bkgd\_Sample,SD5} \sim \text{Normal}(0,1). \\
\sigma \sim \text{Uniform}(0,50)
\end{gathered}
\tag{35}
$$

Parameter values for the prior distribution of μ_α are equal to the parameter values for the α-intercept in the original interaction model. A half-Cauchy distribution was used for the prior distribution of σ_α. It is a typical distribution to use in MCMC sampling models when nothing is assumed about the standard deviation (McElreath, 2016). In this model, the outcome variable $Count$ is a vector of varying intercept parameters. This type of model is typically applied when clustering variables, a method that groups similar variables into representative groups (SAS Institute, 2018). This is actually accomplished with the incorporation of the categorical predictor in the original interaction model, but the purpose of testing this model is to see if a simple multilevel model can improve detection efficiency by creating better parameter estimates. The model was tested using $n = 5$ measurements for 5 different observations. γ_{diff} was calculated

from the estimates and compared to the γ_{diff} calculations from the original interaction model using the same dataset. The comparison is shown in Table 3-10.

Table 3-10. γ_{diff} Estimates for 5 Individual Observations Using the Original Interaction Model and the Multilevel Interaction Model Using MCMC for Various Source Types

200cm		400cm		Bkgd2	
Original	MCMC	Original	MCMC	Original	MCMC
0.857	0.835	0.327	0.347	0.025	0.044
0.982	0.957	0.601	0.576	0.257	0.291
0.656	0.653	0.730	0.707	0.405	0.405
0.379	0.373	0.972	0.934	0.861	0.855
0.228	0.239	0.196	0.205	0.957	0.900

It can be seen that no significant difference exists between the two models regardless of the dataset. The result suggests that for this simpler model, MCMC estimates produce γ_{diff}values similar to the original interaction model and detection efficiencies are expected to be similar. However, this does not mean that more complex models would not perform better than the models presented here. Presumably, given enough knowledge, a Bayesian approach would perfectly model the relationship between $SD5$ and gross counts in a measurement. But, calculating the estimates for these models requires computation time not afforded in the detection scenario presented.

Operational Considerations

A wide range of Bayesian applications to radiation detection exists for source localization and identification and characterization of radioactive samples. These techniques are limited by the requirement of highly specified training datasets and relatively long computation times. While their accuracy and precision is sufficient given enough acquisition time, a Bayesian statistical approach to rapid detection using gross count measurements provides a basic detection method that can be used in low fidelity systems. These systems are typically utilized in

stationary portal monitors. The interaction model presented has been tested under an acquisition time interval equivalent to 5 s on measurements collected by a stationary, continuously operating instrument. The model structure allows for the use of the quadratic approximation to calculate posterior estimates required for detection, a technique that is not computationally demanding and bears a negligible time increase to decision making. The set up for this system would require simultaneous acquisition and analysis of measurements from background and the sample in question. Data collected and believed to be background would serve as the training dataset. This component of the detection system is advantageous in that an established training dataset is not necessary and long run background measurements are not required to establish parameter estimates, unlike the frequentist decision threshold. To ensure that the results observed thus far are not a result of the specific structure of the datasets analyzed, the data were randomized and new $SD5$ were calculated. The new data arrays consisted of the same data used to generate the 1800 individual 1 s measurement, and the data were again partitioned into five 1 s measurement sequences. The results of the analysis are shown in Figures 3-19, 3-20, and 3-21.

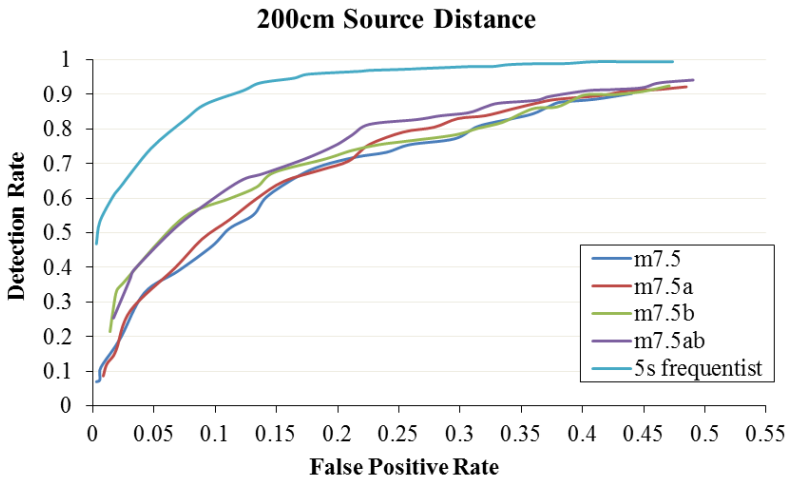

Figure 3-19. Detection efficiencies for the various interaction models and the frequentist decision threshold (5s traditional) for the 200cm source distance using randomized data

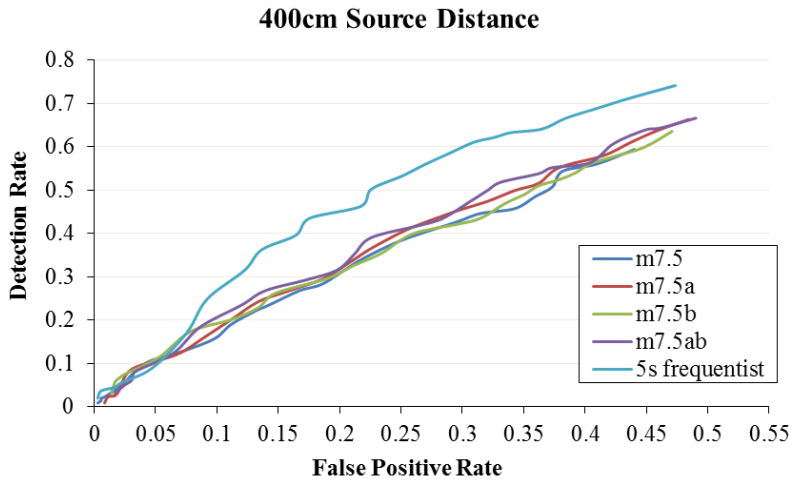

Figure 3-20. Detection efficiencies for the various interaction models and the frequentist decision threshold (5s traditional) for the 400cm source distance using randomized data

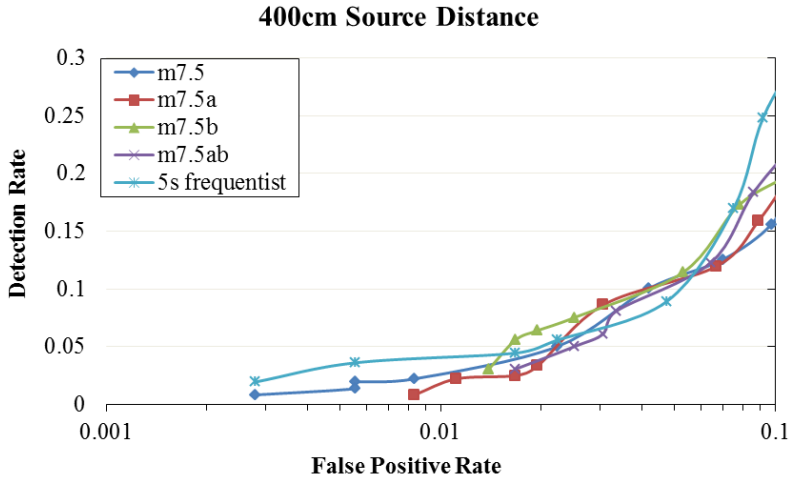

Figure 3-21. Detection efficiencies for the various interaction models and the frequentist decision threshold (5s traditional) for the 400cm source distance for false positives < 0.1 (log scale) using randomized data

Similar trends are observed here when compared to the original dataset, showing that the observed results are not simply due to a particular dataset. The interaction model is, however, extremely sensitive to statistical differences in training dataset distributions. As seen in Table 2-1, the mean of the distributions for Bkgd1 and Bkgd2 are nearly identical, while the mean of the distribution for Bkgd3 is lower than the means of Bkgd1 and Bkgd2 by approximately 2 counts. Figures 3-22 and 3-23 display results from the tests identical to the 5 s measurement experiments

for the 400 cm source distance using the original interaction model presented throughout this dissertation, with the exception that the training dataset is Bkgd2 and the dataset used to test the false positive rate is Bkgd3.

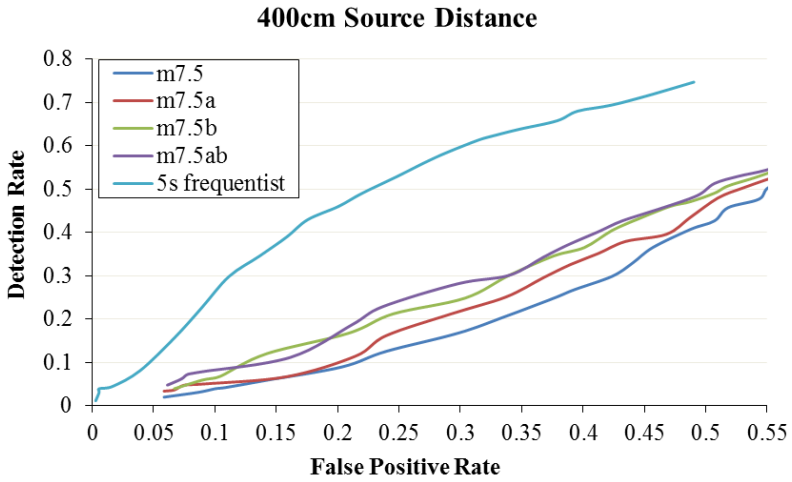

Figure 3-22. Detection efficiencies for the various interaction models and the frequentist decision threshold (5s traditional) for the 400cm source distance when the training dataset and the background dataset differ statistically

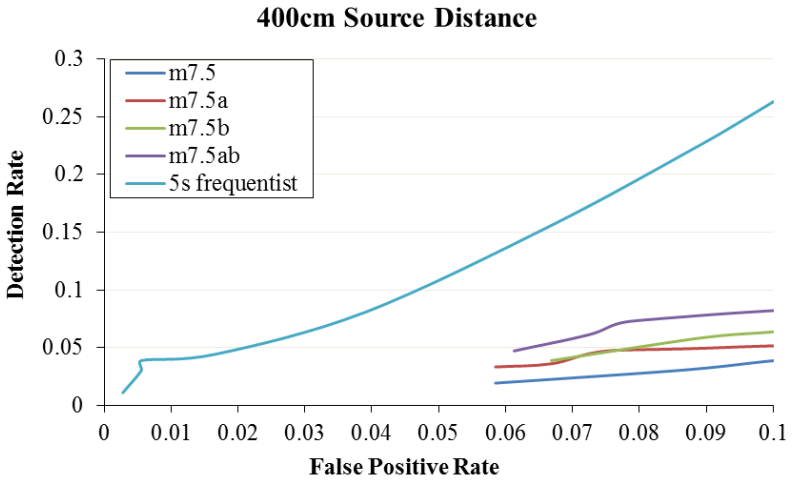

Figure 3-23. Detection efficiencies for the various interaction models and the frequentist decision threshold (5s traditional) for the 400cm source distance when the training dataset and the background dataset differ statistically for false positives < 0.1

The detection efficiency dropped drastically. This is due to a large amount of false positives occurring (triggers on Bkgd3). This result highlights what would happen if the training dataset was acquired in a location where the background differed from the location where the sample in

question was being measured. This is not due to a "unique" order of the dataset, as it was verified in Figures 3-19 through 3-21 that the model is not influenced by the structure of a dataset. This result is extremely important to consider given that all of the tests presented were carried out on data where the source was stationary and continuously within the field of the window of the detector. In a scenario where the source is not stationary, the gross count measurement per time interval is roughly a function of distance from the detector window. The interaction model would be sensitive to the changes occurring per unit time interval, especially one in which the source is passing by the detector. The same datasets used to create Figures 3-22 and 3-23 were applied to the flipped categorical predictor model to verify if the same sensitivity occurred. Figures 3-24 and 3-25 display the results.

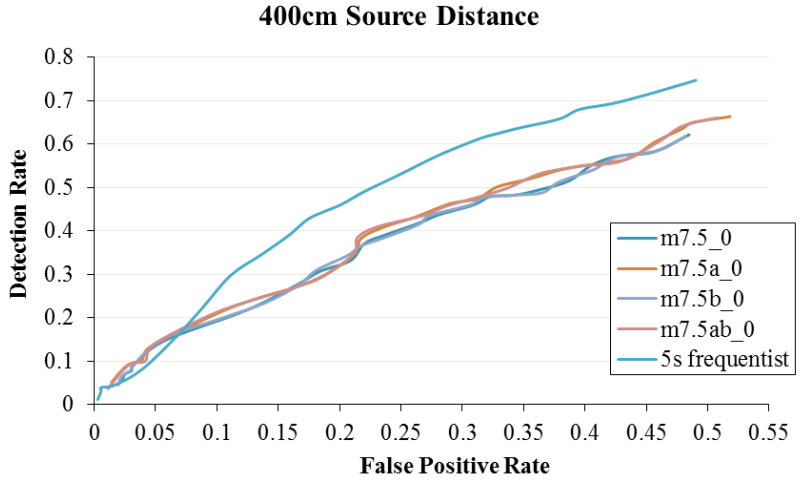

Figure 3-24. Detection efficiencies for the various flipped categorical predictor interaction models and the frequentist decision threshold (5s traditional) for the 400cm source distance when the training dataset and the background dataset differ statistically

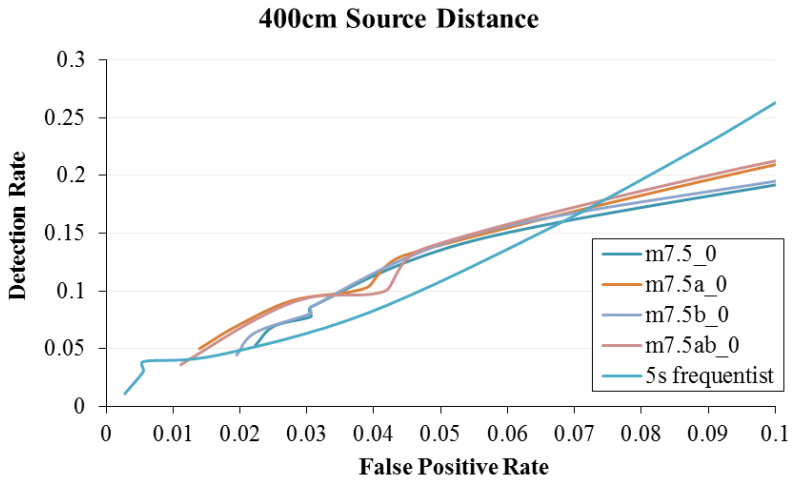

Figure 3-24. Detection efficiencies for the various flipped categorical predictor interaction models and the frequentist decision threshold (5s traditional) for the 400cm source distance when the training dataset and the background dataset differ statistically for false positives < 0.1

These results look similar to the models where statistically similar background distributions were used. Given that the flipped categorical predictor model "washed out" the sensitivity of the interaction model, it may be best to only use interaction models where the categorical predictor assigns $Bkgd\_Sample = 1$ if $Count_i$ is attributed to a background measurement; 0 otherwise.

CONCLUSION

Numerous studies have been published using Bayesian statistics in source localization and identification, characterization of radioactive samples, and uncertainty analysis; but there is a limited amount of material specific to the development of a decision threshold for simple gross count measurements using Bayesian statistics. Difficulties arise when applying decision techniques to low count rate data, which are restricted by the fact that decisions are being made on individual gross count measurements alone. As the overlap of the distributions of the background and source increases, the number of alarms triggered with a source present decreases. A decision threshold on individual measurements will always be limited by this caveat. The investigation presented demonstrates a method to develop a viable Bayesian model to detect radiological sources using gross count measurements in low fidelity systems. Bayesian statistics provides a theoretically infinite number of approaches to this scenario; the challenge lies in creating an approach that is operationally tractable while adhering to Bayesian modeling strategies.

The Bayesian interaction model was proven statistically by comparing a multivariate regression model with and without interaction to study if the relationship between $SD5$ and gross counts in a measurement depends upon whether or not a source is present. The interaction model was then verified through the fact that these types of models are symmetrical and finding that the relationship between whether or not a source is present and gross counts in a measurement is dependent upon $SD5$. After finding that the interaction model was both valid, and that these relationships existed for the data tested, Bayesian statistical analyses were used to understand the parameters surrounding the original relationship that $SD5$ and gross counts in a measurement

depends upon whether or not a source is present. This relationship is described statistically by the parameter γ; and a perspective was presented, γ_{diff}, that allowed the question "what is the probability that the relationship between $SD5$ and gross count measurements from background is less than the relationship between $SD5$ and gross count measurements from a sample?" This question served as the framework for a decision rule, $\Pr^* < \gamma_{diff_i}$, which can be used in detection scenarios and applied to varying forms of the interaction model. The interaction model was tested on data and compared to a frequentist decision threshold to determine its detection efficiency. The model was first tested to show that the decision rule could be applied to an operationally relevant number of measurements ($n = 5$), and then tested on a larger dataset comprised of this measurement length. The model was found to perform comparably to the 5s frequentist decision threshold for weak sources at lower false positive rates.

Various forms of the model were then developed based on information criterion and tested in the same manner. The scope of the models tested was limited to reasonable changes to parameter estimates established by the maximum entropy principle and occupational limitations to detection due to computation time. WAIC testing did not provide a conclusive model for the best predictive efficiency. Model variations in parameter estimates, flipping the categorical predictor, and changing the nature of the relationship from a linear to a power function displayed similar detection efficiencies. A multilevel model was developed and tested, but it was found to produce γ_{diff} values similar to the original interaction model. Any further model validation using multilevel models was deemed unnecessary due to the computation time required for the MCMC calculations used. These results all suggest that the original linear model is sufficient at this time, and that more rigorous modeling techniques would be required for any possible improvement in detection efficiency.

The interaction model operates by examining a set of previously recorded background gross count measurements, the training data Bkgd1, and resulting $SD5$ with a set of unknown sample gross count measurements and resulting $SD5$. These two sets of data make up the arrays for $Count_i$ and $SD5_i$. Included in the data array is $Bkgd\_Sample_i$, such that each index is correctly categorized as a known background measurement or an unknown sample measurement. This setup allows the model to work in a way that is intuitive to the operational measurement technique: known background data and resulting estimates are used to create a relationship that is expected to be consistent across all measurements with no source present, and this relationship is compared to the samples in question to judge if a source is present. The set up for this system would require simultaneous acquisition and analysis of measurements from background and the sample in question. Data collected and believed to be background would serve as the training dataset. This component of the detection system is advantageous in that an established training dataset is not necessary and long run background measurements are not required to establish parameter estimates, unlike the frequentist decision threshold. The interaction model is extremely sensitive to statistical differences in training dataset distributions. This result is extremely important to consider given that all of the tests presented were carried out on data where the source was stationary and continuously within the field of view of the detector. In a scenario where the source is not stationary, the gross count measurement per time interval is roughly a function of distance from the active region of the detector. The interaction model would be sensitive to the changes occurring per unit time interval, especially one in which the source is passing by the detector. A feature such as this in a string of measurements may not be detectable in the 5 s equivalent of the frequentist decision threshold. This scenario should be tested in a laboratory setting to determine the sensitivity of the interaction model. A larger

dataset than the one presented here should also be tested to study the detective efficiency of the models at extremely low (false positives < 0.01) false positive rates.

A final consideration for the interaction model is that γ, and consequently this categorical linear interaction model, is universal. Even though the scope of the paper studies the relationship between $SD5$ and gross counts in a measurement, in theory any predictor can be used in place of $SD5$. For example, interaction models were developed using $SD4$ (standard deviation of the current measurement and the previous three measurements) and $SD3$ (standard deviation of the current measurement and the previous two measurements) and tested over a small but identical dataset to the ones presented in this dissertation. The regression plots tested against the 400 cm dataset for the interaction model only are shown in Figures 4-1 and 4-2.

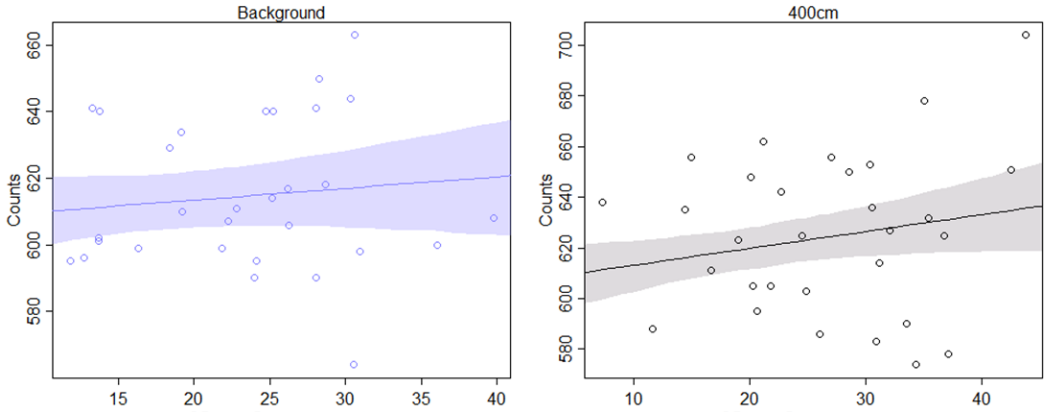

Figure 4-1. Regression plots using $SD4$

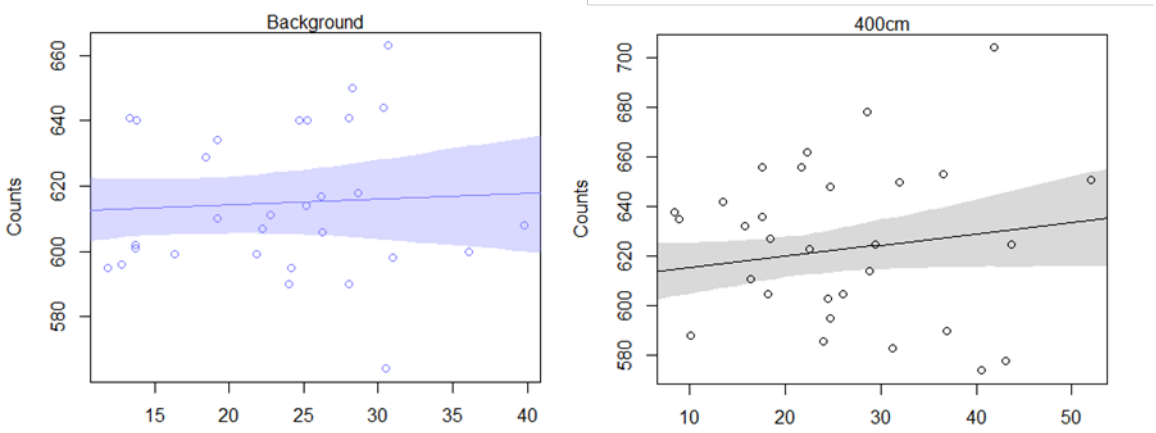

Figure 4-1. Regression plots using $SD3$

The figures show that the interaction model can still detect differences for weak sources using different predictors. This feature of the interaction model is operationally advantageous and highlights the strength of Bayesian statistics. In theory, the experimenter (or health physicist) can construct a parameter they deem important in the detection scenario and use posterior distributions as a means to determine whether or not a source is present. The Bayesian model can provide abstract statistical knowledge about unique parameters, allowing the applications and capabilities of this technique to far outreach those of the frequentist decision threshold applied to gross count measurements.

REFERENCES

Altschuler, B. & Pasternack, B., 1963. Statistical measures of the lower limit of detection of a radioactive counter. *Health Physics,* Volume 9, pp. 293-298.

ANSI N13.30 Rev.1, 2011. *Performance Criterion for Radiobioassay,* s.l.: Health Physics Society.

Berger, J., 1985. *Statistical Decision Theory and Bayesian Statistics.* s.l.:Springer.

Brandl, A., 2013. Statistical considerations for improved signal identification from repeated measurements at low signal-to-background ratios. *Health Physics,* 104(3), pp. 256-263.

Brandl, A. & Herrera Jimenez, A., 2008. Statistical criteria to set alarm levels for continuous measurements of ground contamination. *Health Physics,* 95(Suppl 2), pp. S128-S132.

Brogan, J. & Brandl, A., in press. Enhancing test statistics by utilizing data patters in sequential measurement strings in radiation detection. *Health Physics.*

Cember, H. & Johnson, T., 2009. *Introduction to Health Physics.* 4 ed. s.l.:McGraw-Hill.

Chipman, H., George, E. & McCulloch, R., 2001. Practical Implementation of Bayesian Model Selection. *IMS Lecture Notes - Monograph Series,* Volume 38.

Chistenson, R., Johnson, W., Branscum, A. & Hanson, T., 2011. *Bayesian Ideas and Data Analysis.* Boca Raton, FL: CRC Press.

Currie, L., 1968. Limits for Qualitative Detection and Quantitative Determination. *Analytical Chemistry,* 40(3), pp. 586-593.

Gelman, A. et al., 2013. *Bayesian Data Analysis.* 3rd ed. Boca Raton, FL: CRC Press.

Gelman, A. & Shalizi, C., 2013. Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology,* Volume 66, pp. 8-38.

International Organization for Standardization, 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.

IUPAC, 1995. Nomenclature in Evaluation of Analytical Methods Including Detection and Quantification Capabilities. *Pure & Appl. Chem.,* 67(10), pp. 1699-1723.

Klumpp, J., Miller, G. & Poudel, D., 2018. A new approach to counting measurements: Addressing the problems with ISO-11929. *Nuclear Inst. and Methods in Physics Research,* A(892), pp. 18-29.

Kruschke, J., 2010. Doing Bayesian data analysis: A tutorial introduction to R. *Academic Press,* pp. 88-90.

MARLAP, 2004. Multi-Agency Radiological Laboratory Analytical Protocols Manual. Volume III, pp. Chapters 18-20 and Appendix G.

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

Michel, R., 2016. Measuring, Estimating, and Deciding under Uncertainty. *Applied Radiation and Isotopes,* Volume 109, pp. 6-11.

National Council on Radiation Protection and Measurements, 1985. *A handbook of radioactivity measurements procedures. 2nd ed.,* Bethesda, MD: NCRP Publications; NCRP Report 58.

NumFOCUS, 2018. *Stan.* [Online]
Available at: http://mc-stan.org/

NUREG-4007, 1984. *Lower Limit of Detection: Definition and Elaboration of a Proposed Position for Radiological Effluent and Environmental Measurements,* Washington, DC: National Bureau of Standards.

Rucker, T. L., 2001. Calculation of decision levels and minimum detectable concentrations from method blank and sample uncertainty data - utopian statistics. *Journal of Radioanalytical and Nuclear Chemistry,* 248(1), pp. 191-196.

SAS Institute, 2018. *Clustering Variables.* [Online]
Available at: https://www.jmp.com/support/help/14/cluster-variables.shtml
[Accessed 1 May 2018].

Silver, N., 2012. *The signal and the noise.* New York, NY: Penguin Press.

Strom, D. & MacLellan, J., 2001. Evaluation of Eight Decision Rules for Low-Level Radioactivity Counting. *Health Physics,* 81(1), pp. 27-34.

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

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

Turner, J. E., 2007. *Atoms, Radiation, and Radiation Protection.* 3rd ed. Oak Ridge, TN: Wiley-VCH.

Vallverdu, J., 2008. The false dilemma: Bayesian vs. frequentist. *arXiv: 0804.0486v1.*

Wackerly, D., Mendenhall, W. & Scheaffer, R., 2008. *Mathematical Statistics with Applications.* 7th ed. Belmont, CA: Brooks/Cole CENGAGE Learning.

Wald, A., 1947. An essentially complete class of admissible decision functions. *The Annals of Mathematical Statistics,* 18(4), pp. 549-555.

APPENDIX A

Regression and Symmetry Plots Using Bkgd2 as the Training Dataset

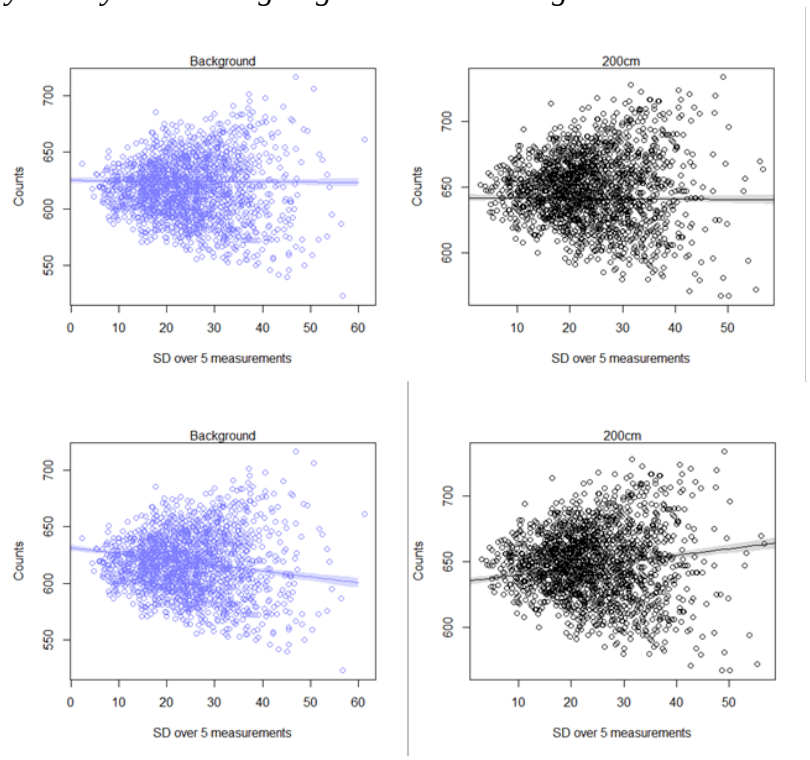

Figure A-1. Regression plots for the 200cm source distance

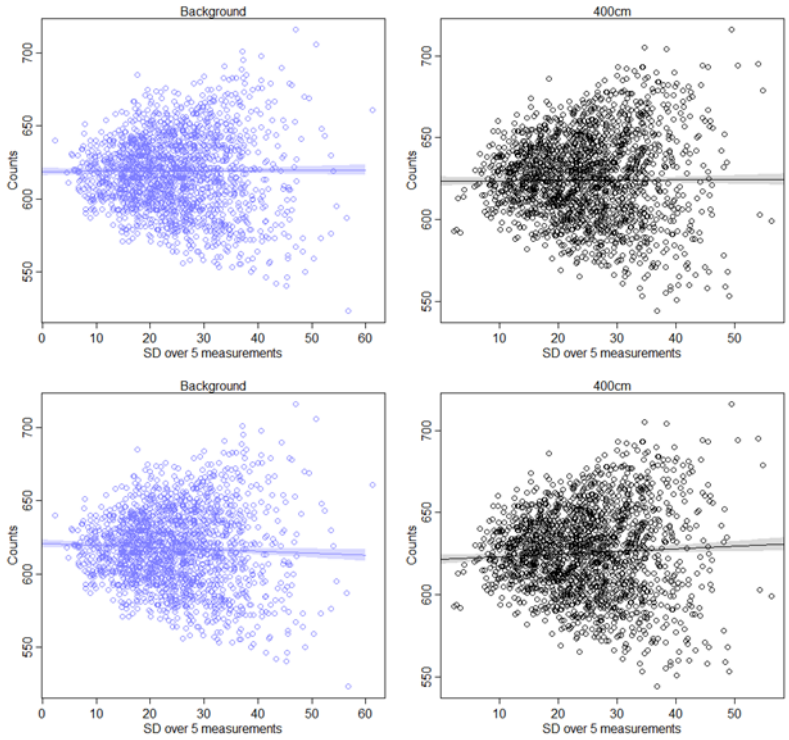

Figure A-2. Regression plots for the 400cm source distance

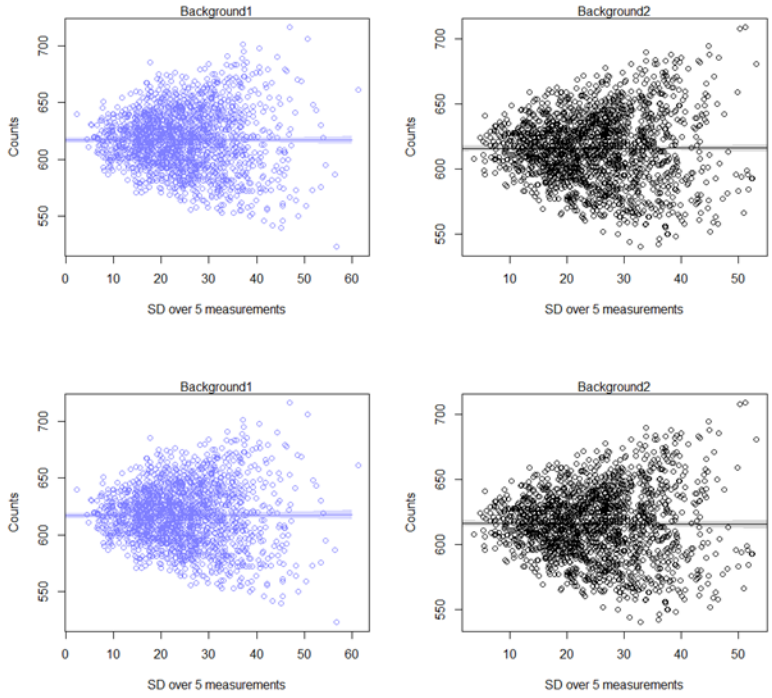

Figure A-3. Regression plots for the Bkgd3 source distance. Note that Background1 is actually the Bkgd2 dataset, and Background2 is the Bkgd3 dataset

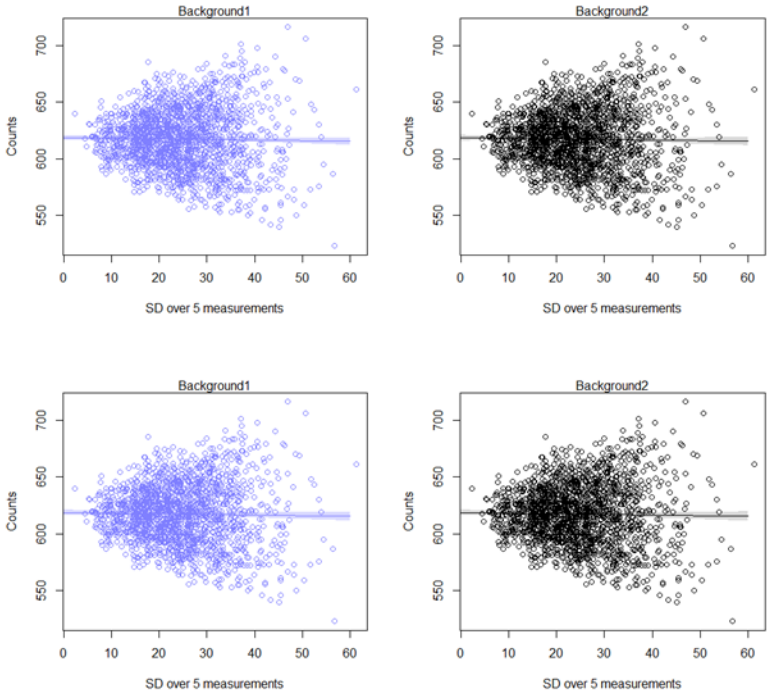

Figure A-4. Regression plots for the Self-test

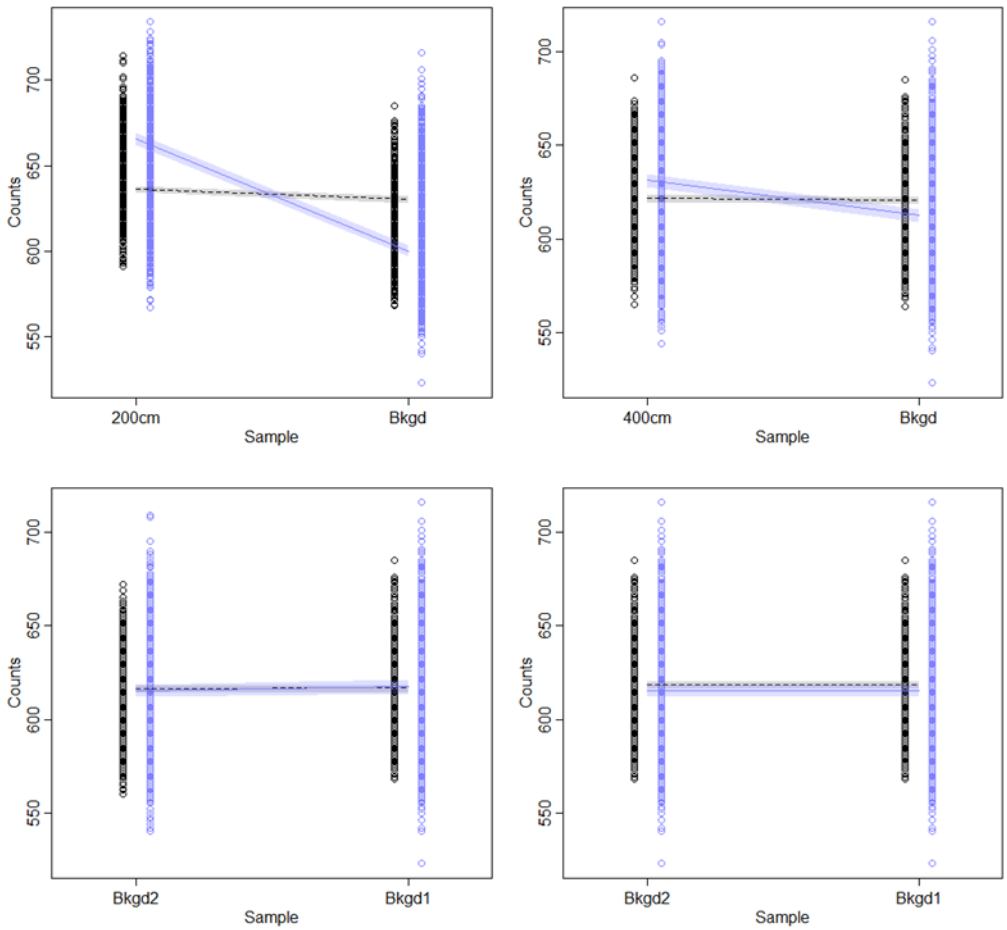

Figure A-5. Symmetry plots comparing Bkgd2 (labeled Bkgd/Bkgd1) to the 200cm source distance (top left) dataset, the 400cm source distance dataset (top right), the Bkgd3 dataset (Bkgd2 in bottom left), and the Self-test (bottom right)

Regression and Symmetry Plots Using Bkgd3 as the Training Dataset

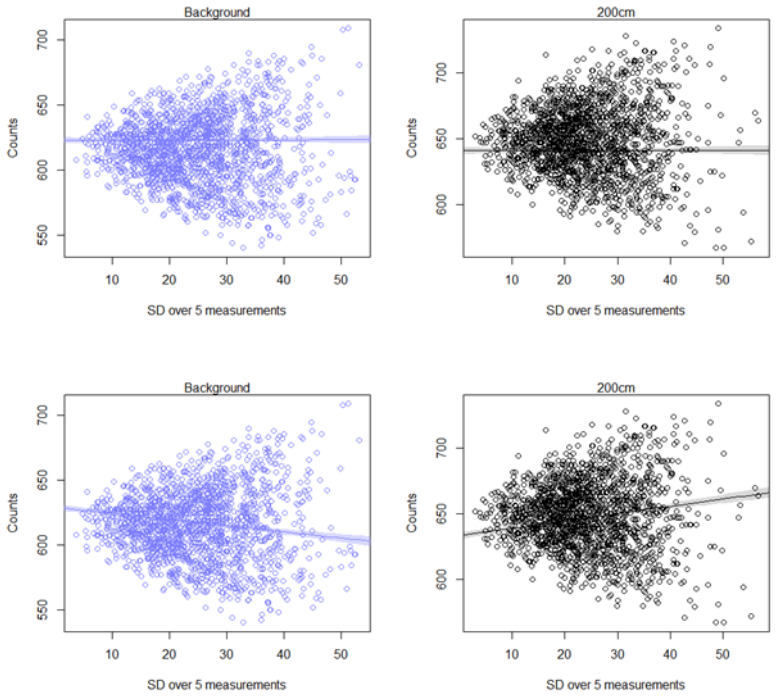

Figure A-6. Regression plots for the 200cm source distance

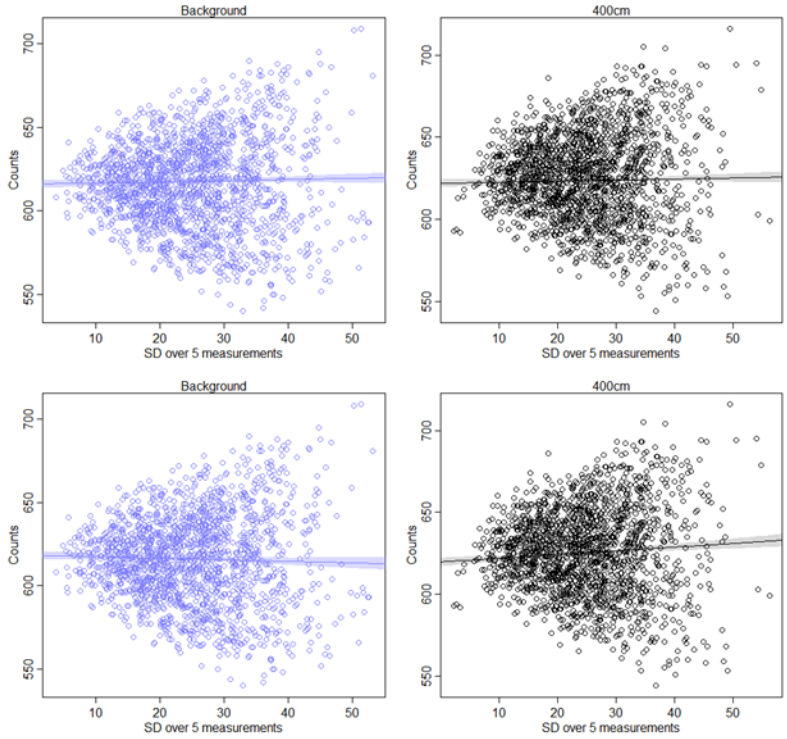

Figure A-7. Regression plots for the 400cm source distance

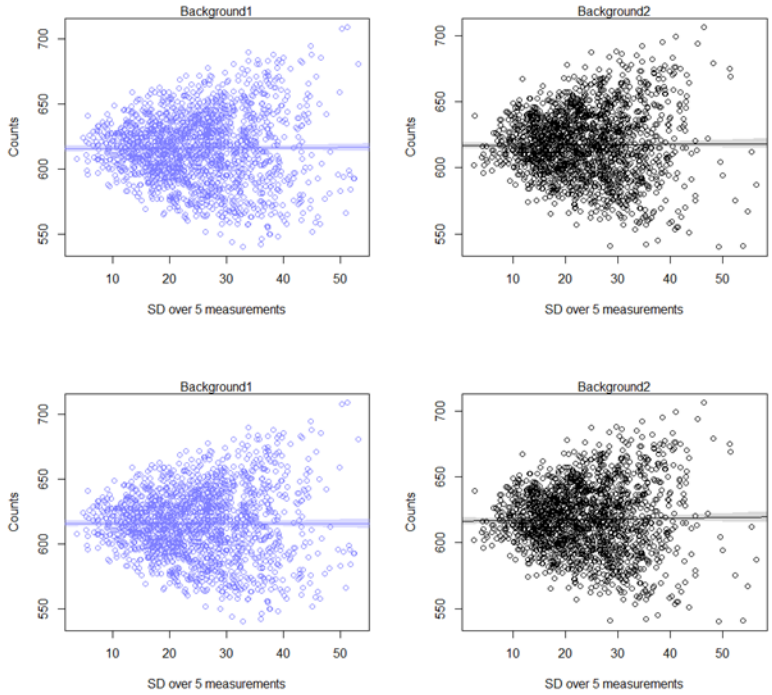

Figure A-8. Regression plots for the Bkgd1 source distance. Note that Background1 is actually the Bkgd3 dataset, and Background2 is the Bkgd1 dataset

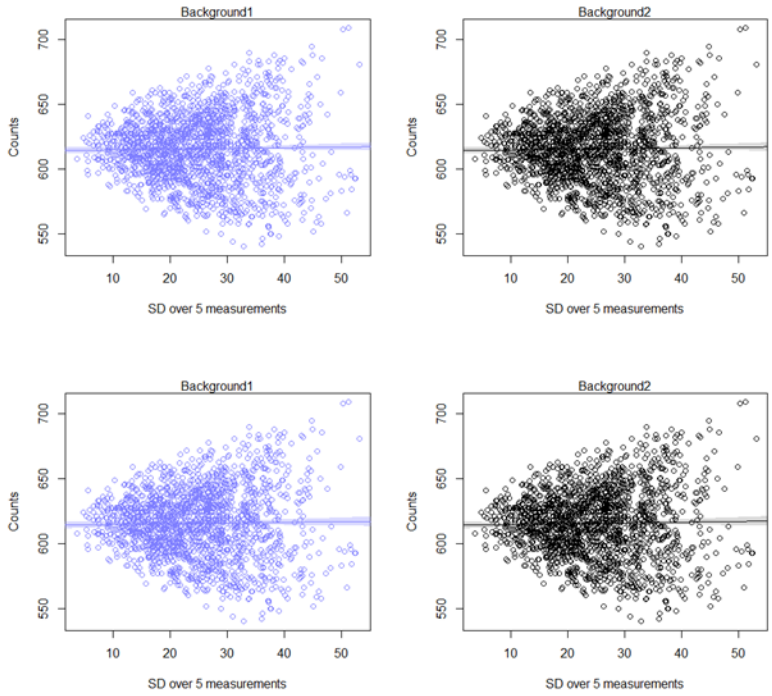

Figure A-9. Regression plots for the Self-test

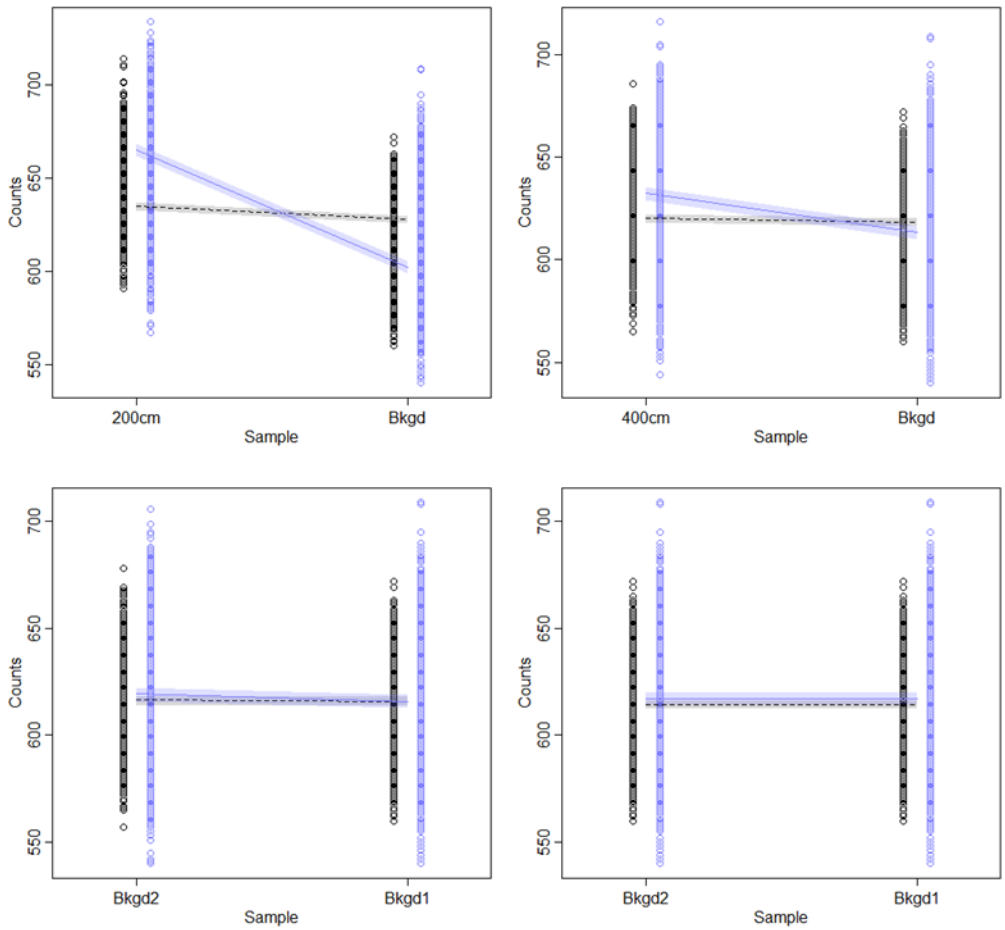

Figure A-10. Symmetry plots comparing Bkgd2 (labeled Bkgd/Bkgd1) to the 200cm source distance (top left) dataset, the 400cm source distance dataset (top right), the Bkgd3 dataset (Bkgd2 in bottom left), and the Self-test (bottom right)

γ_{diff} Distributions

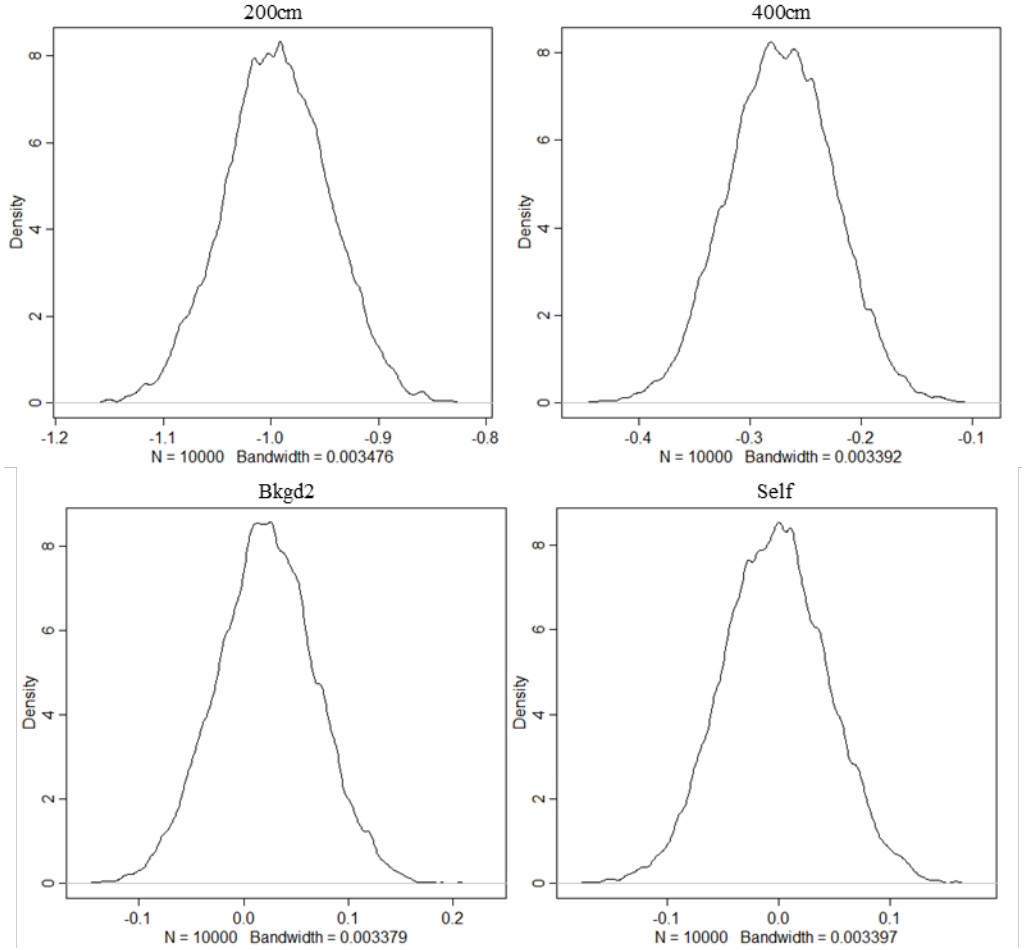

Figure A-11. Distributions of γ_{diff} for labeled source distance datasets using the Bkgd1 training dataset

APPENDIX B

$n = 900$ *Data*

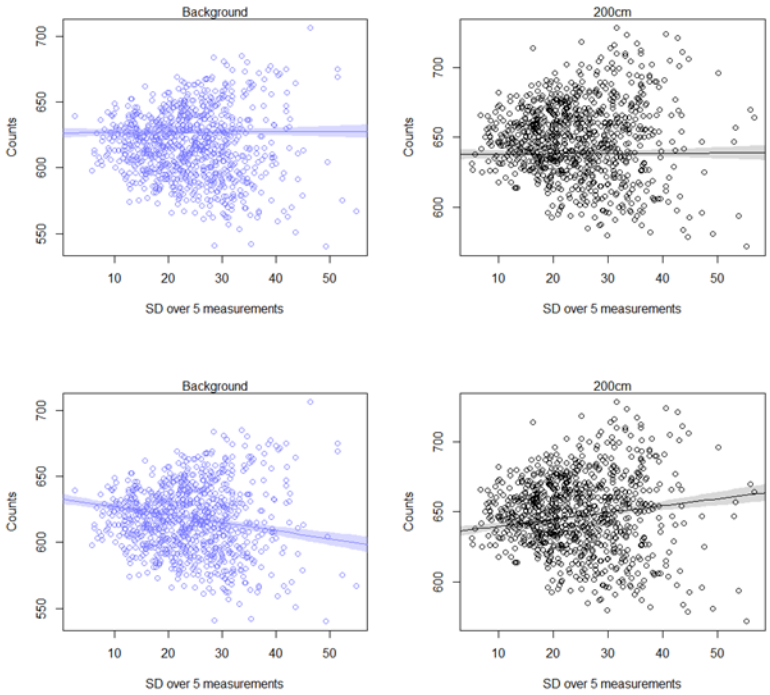

Figure B-1. Regression plots for the 200cm source distance

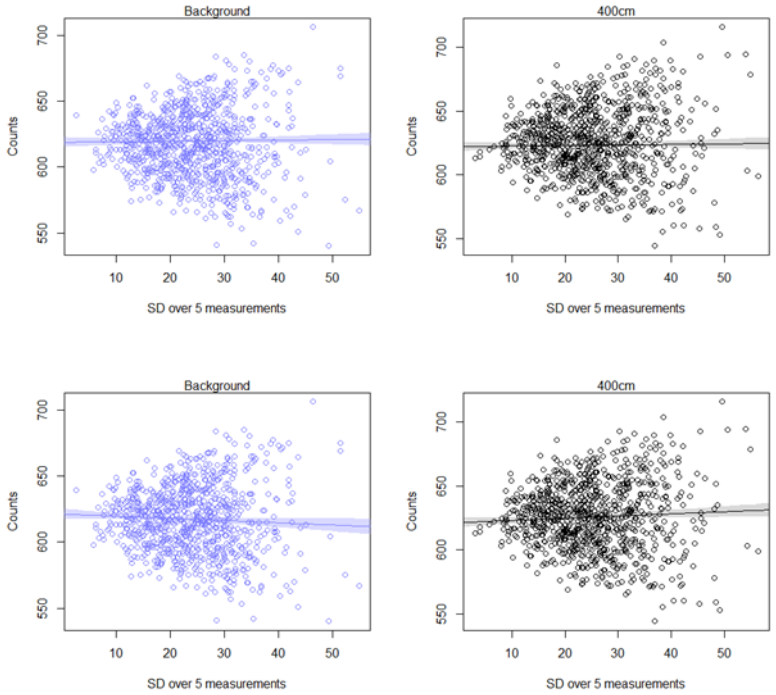

Figure B-2. Regression plots for the 400cm source distance

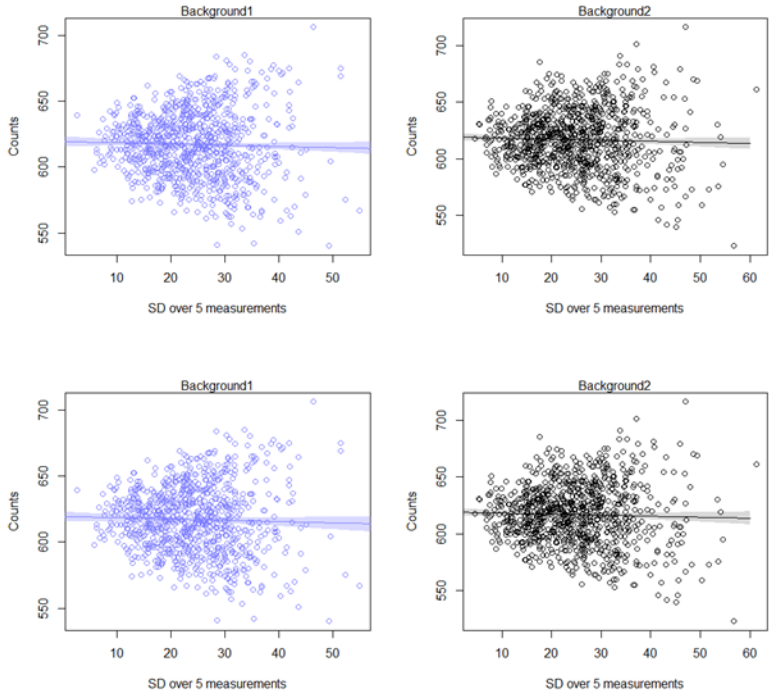

Figure B-3. Regression plots for the Bkgd2 source distance

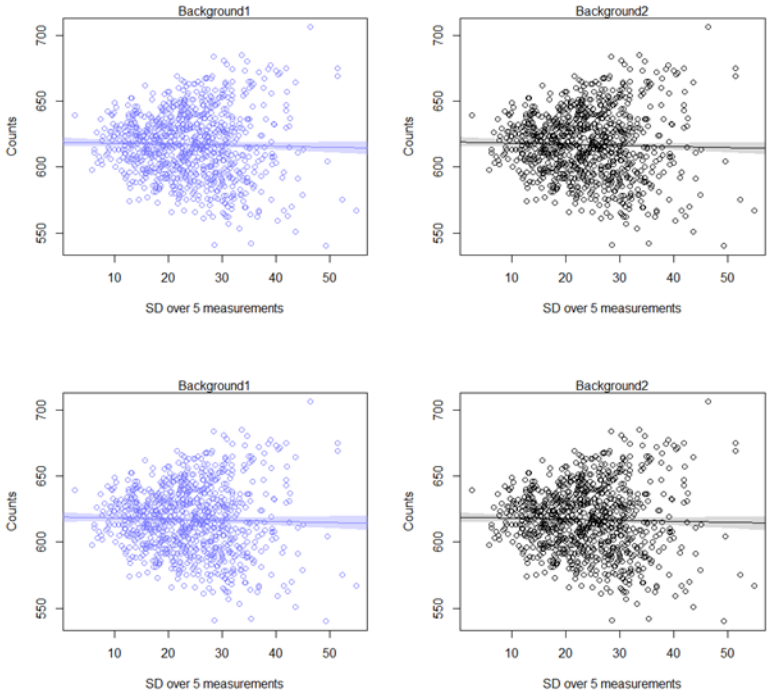

Figure B-4. Regression plots for the Self-test

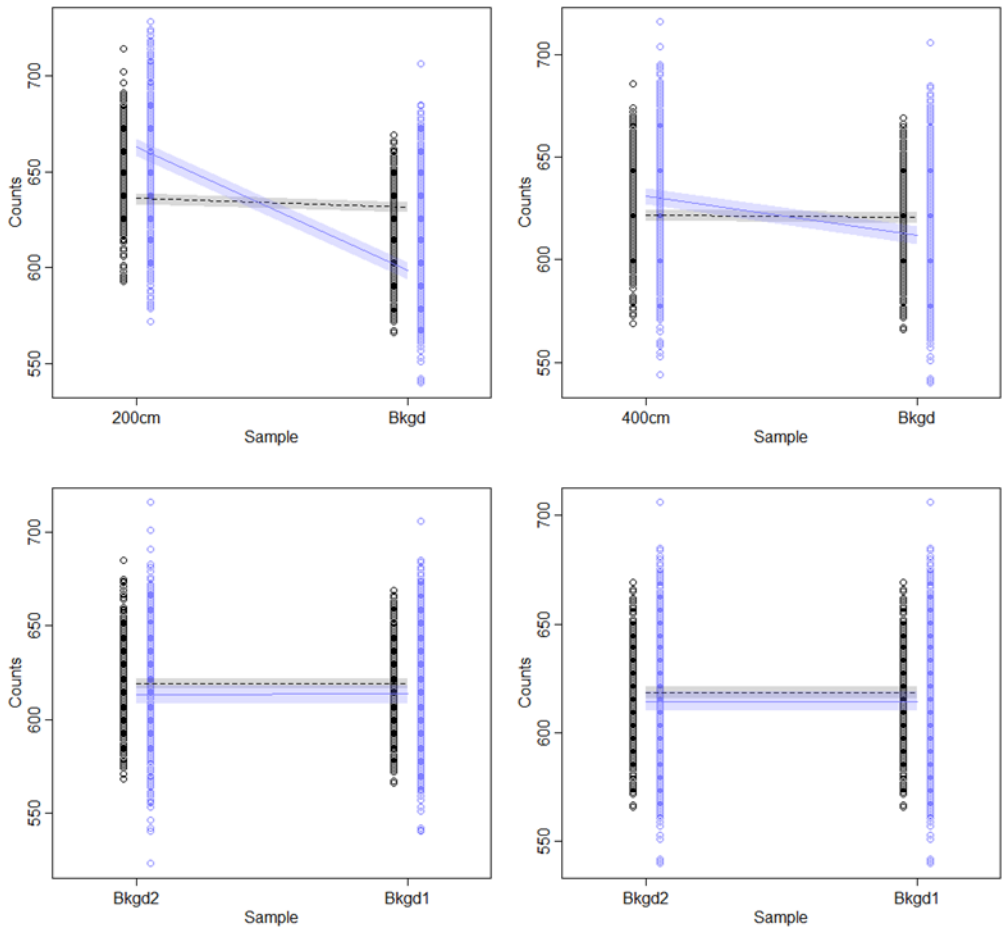

Figure B-5. Symmetry plots comparing Bkgd1 (labeled Bkgd/Bkgd1) to the 200cm source distance (top left) dataset, the 400cm source distance dataset (top right), the Bkgd2 dataset (Bkgd2 in bottom left), and the Self-test (bottom right)

Table B-1. MAP Estimates for Parameters from the Interaction Model

	Unknown Sample Dataset							
	200cm		400cm		Background 2		Self	
parameter θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ
α	634.92	1.9	621.37	1.87	619.15	1.83	618.8	1.86
β_{SD5}	0.49	0.08	0.17	0.07	-0.09	0.07	-0.08	0.08
β_{Bkgd_Sample}	-1.9	0.97	-0.25	0.96	0.04	0.96	0.09	0.96
$\beta_{Bkgd\_Sample,SD5}$	-1.1	0.06	-0.34	0.06	0	0.06	0	0.06
σ	26.33	0.44	25.89	0.43	25.96	0.43	25.38	0.42

Table B-2. Calculated γ for Given Source Types

Source Type	γ	
	No Source	Source
200cm	-0.61	0.49
400cm	-0.16	0.17
Background 2	-0.09	-0.09
Self	-0.08	-0.08

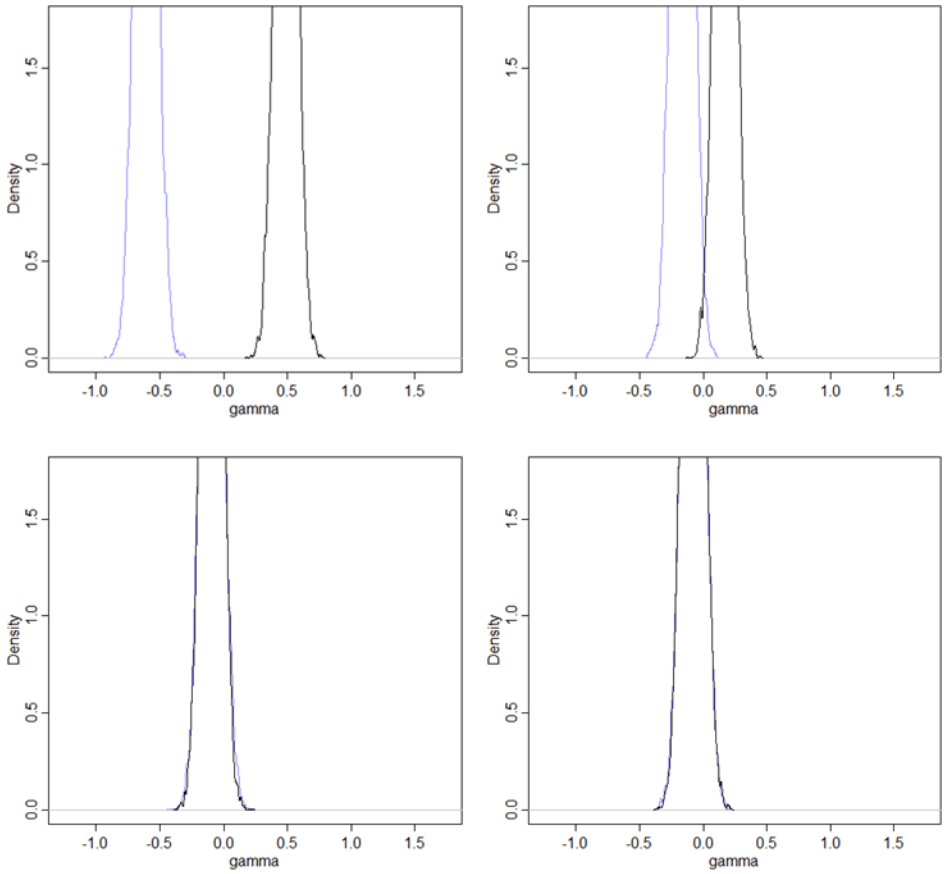

Figure B-6. Comparisons of marginal distributions for Bkgd1 (violet) versus source type 400cm (top left), 400cm (top right), Bkgd2 (bottom left), and Self (bottom right)

Table B-3. Means of the Marginal Distributions for a Given Source Type

Source Type	μ_γ	
	No Source	Source
200cm	-0.6082	0.4876
400cm	-0.1636	0.1731
Background 2	-0.0919	-0.0934
Self	-0.0786	-0.0766

$n = 450$ *Data*

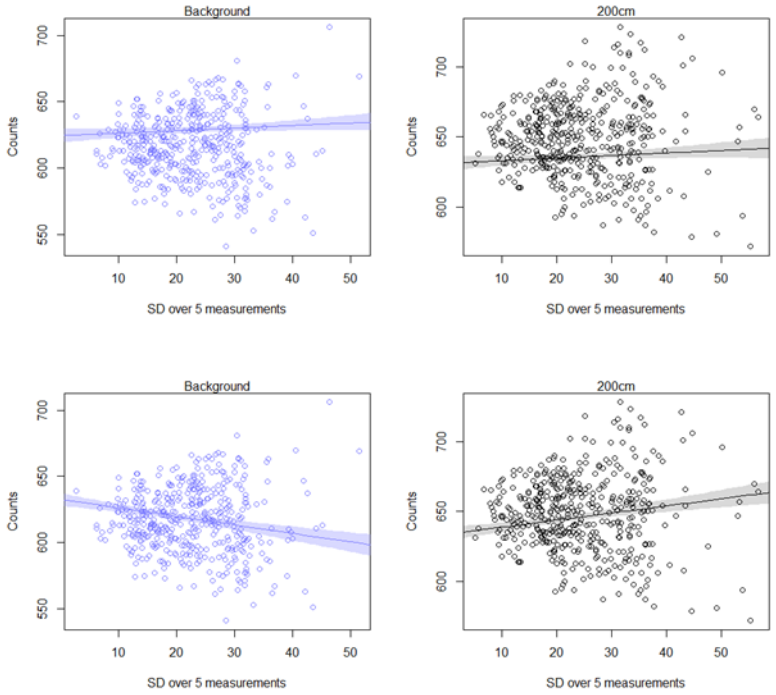

Figure B-7. Regression plots for the 200cm source distance

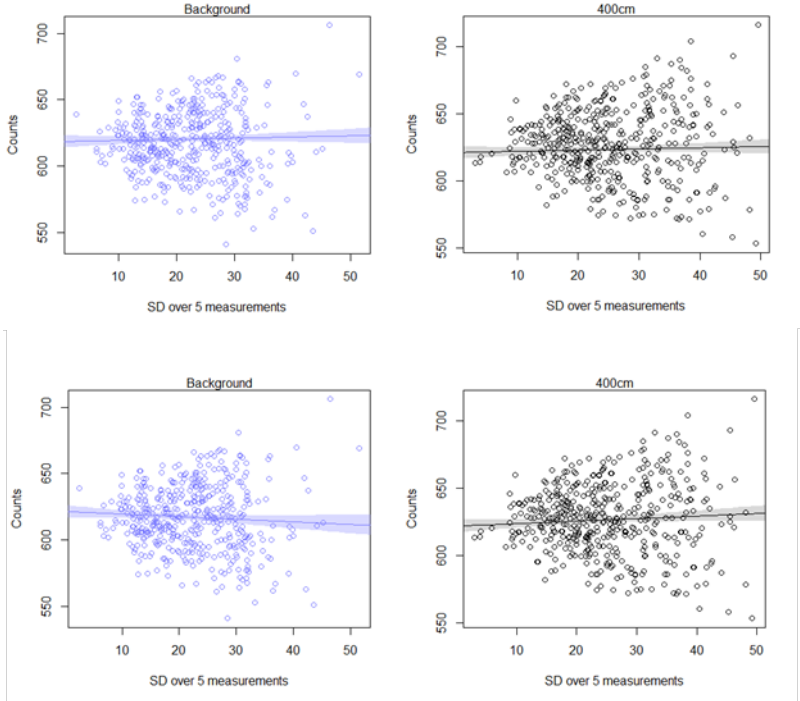

Figure B-8. Regression plots for the 400cm source distance

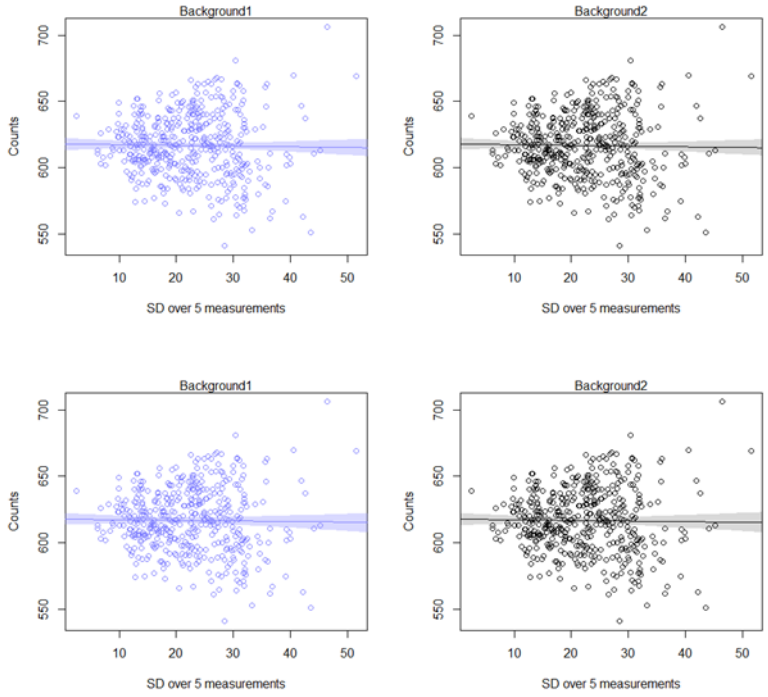

Figure B-9. Regression plots for the Bkgd2 source distance

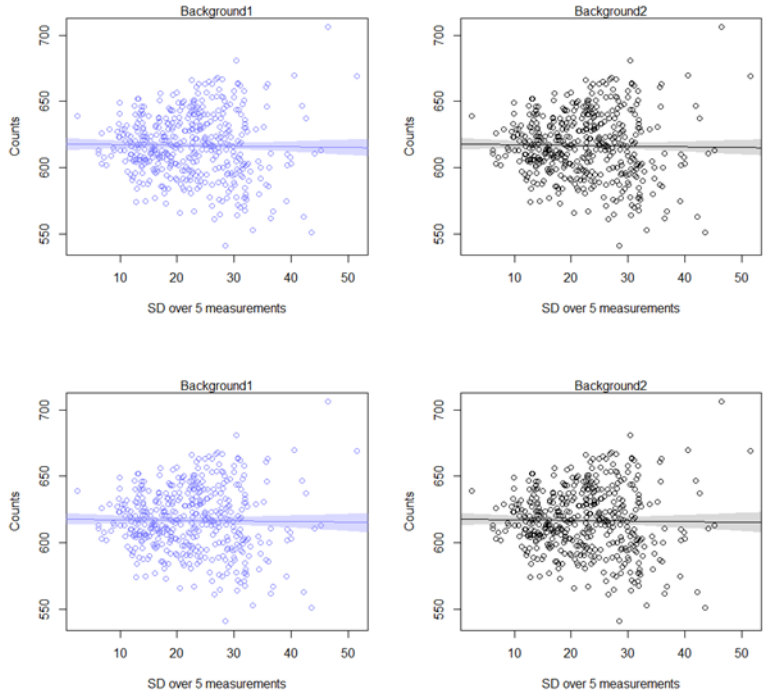

Figure B-10. Regression plots for the Self-test

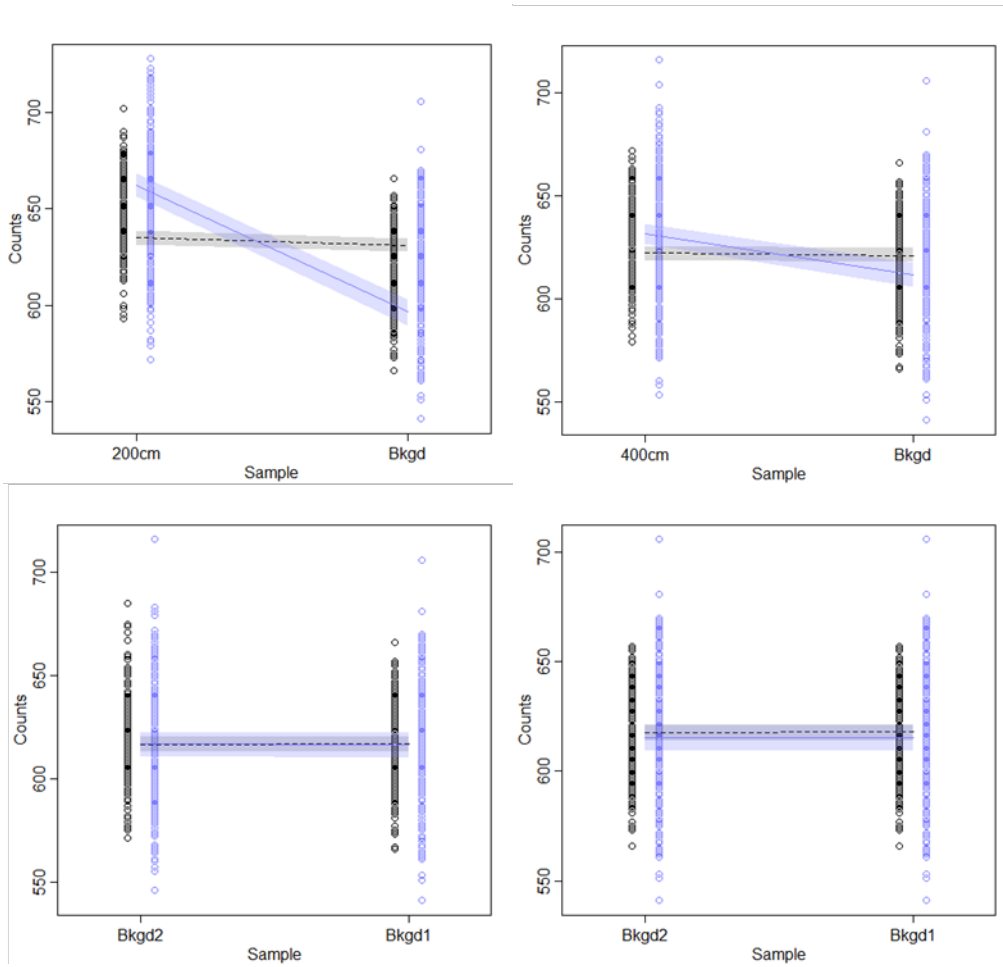

Figure B-11. Symmetry plots comparing Bkgd1 (labeled Bkgd/Bkgd1) to the 200cm source distance (top left) dataset, the 400cm source distance dataset (top right), the Bkgd2 dataset (Bkgd2 in bottom left), and the Self-test (bottom right)

Table B-4. MAP Estimates for Parameters from the Interaction Model

	Unknown Sample Dataset							
	200cm		400cm		Background 2		Self	
parameter θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ
α	633.78	2.48	621.7	2.45	616.62	2.47	617.83	2.4
β_{SD5}	0.51	0.1	0.19	0.1	0.01	0.1	-0.05	0.11
β_{Bkgd_Sample}	-1.06	0.98	-0.21	0.98	0.17	0.98	0.09	0.98
$\beta_{Bkgd\_Sample,SD5}$	-1.14	0.08	-0.38	0.08	-0.01	0.08	0	0.08
σ	25.99	0.61	25.23	0.6	25.42	0.6	24.26	0.57

Table B-5. Calculated γ for Given Source Types

Source Type	γ	
	No Source	Source
200cm	-0.61	0.49
400cm	-0.19	0.19
Background 2	-0.09	-0.09
Self	-0.05	-0.05

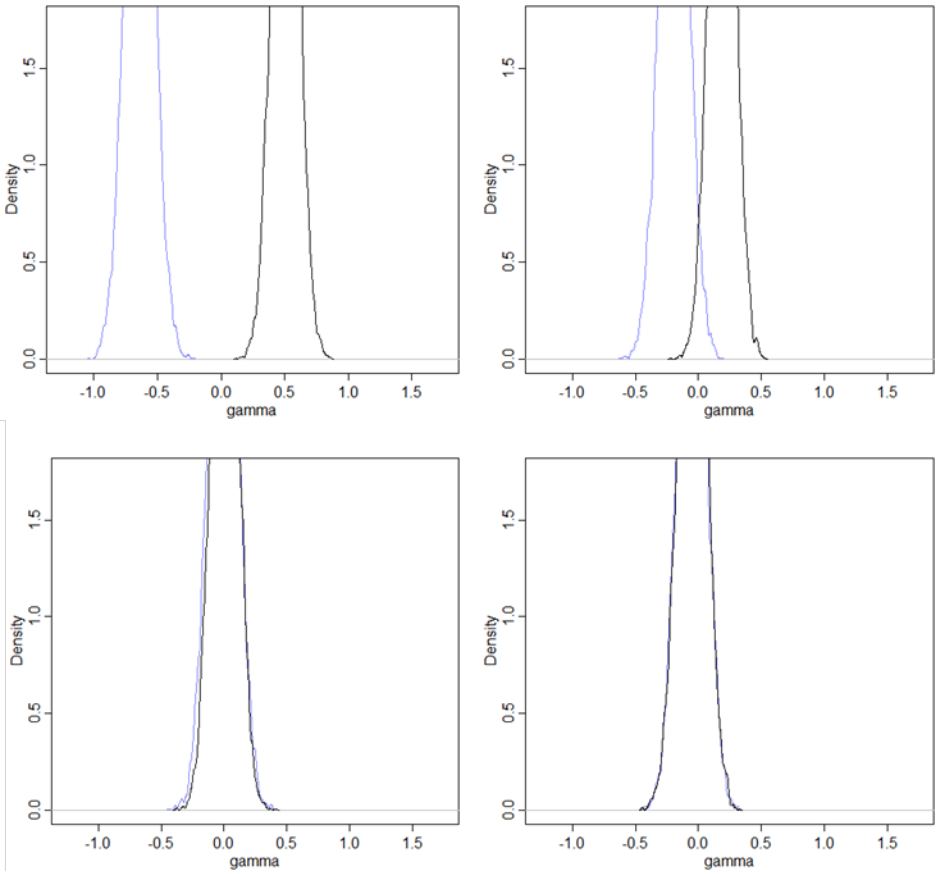

Figure B-12. Comparisons of marginal distributions for Bkgd1 (violet) versus source type 400cm (top left), 400cm (top right), Bkgd2 (bottom left), and Self (bottom right)

Table B-6. Means of the Marginal Distributions for a Given Source Type

Source Type	μ_γ	
	No Source	Source
200cm	-0.6396	0.5042
400cm	-0.194	0.1871
Background 2	-0.0072	0.0043
Self	-0.0521	-0.0493

$n = 225$ *Data*

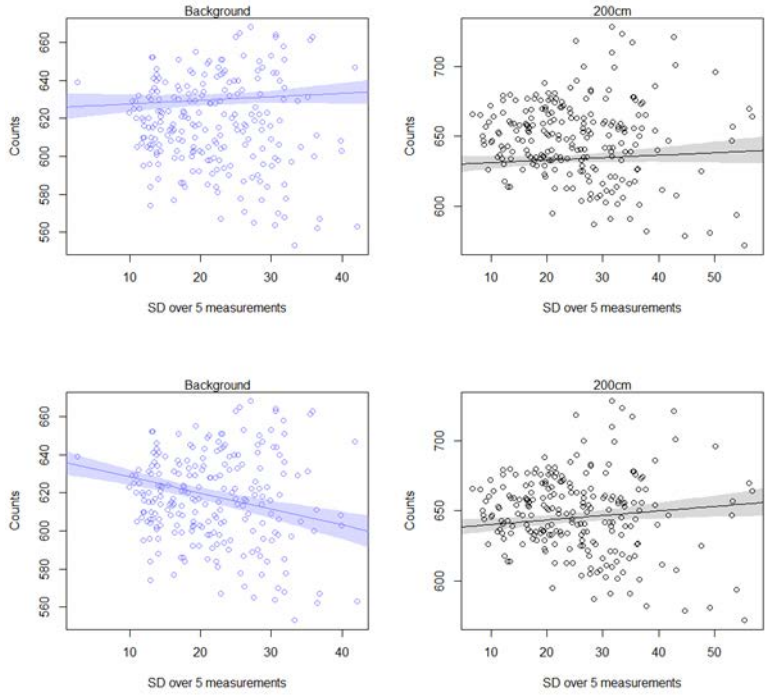

Figure B-13. Regression plots for the 200cm source distance

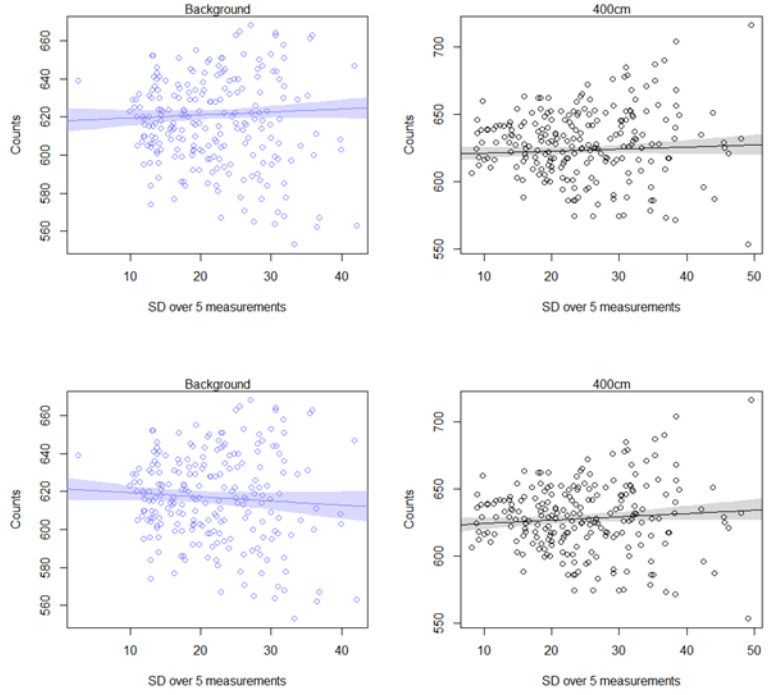

Figure B-14. Regression plots for the 400cm source distance

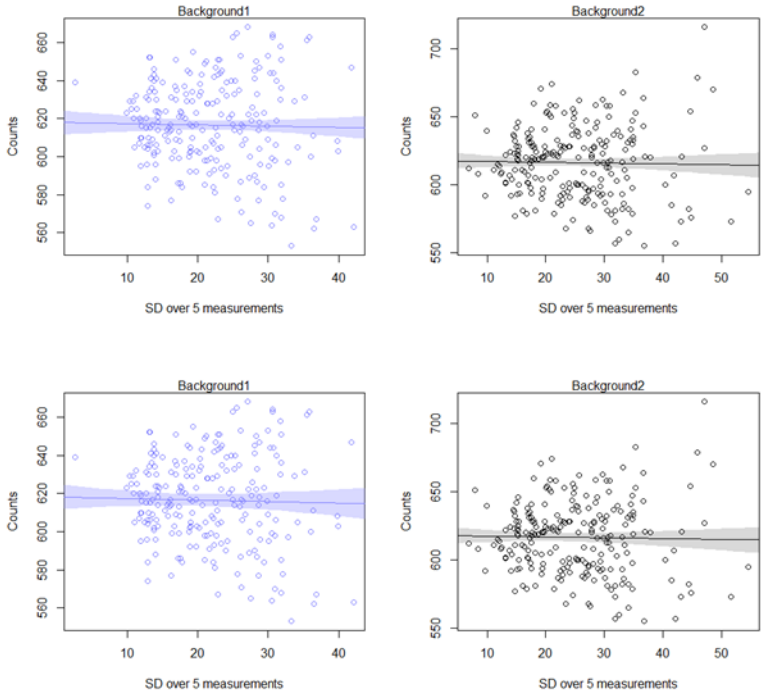

Figure B-15. Regression plots for the Bkgd2 source distance

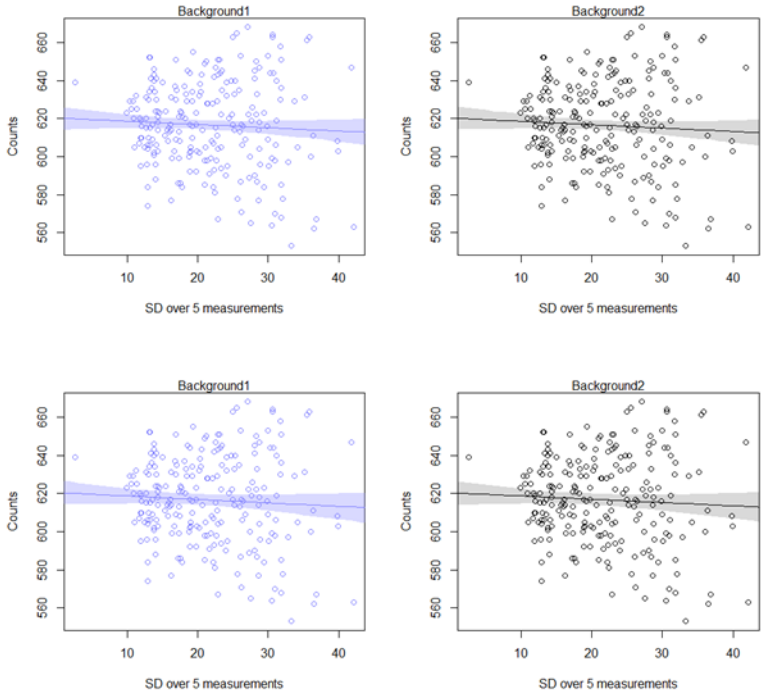

Figure B-16. Regression plots for the Self-test

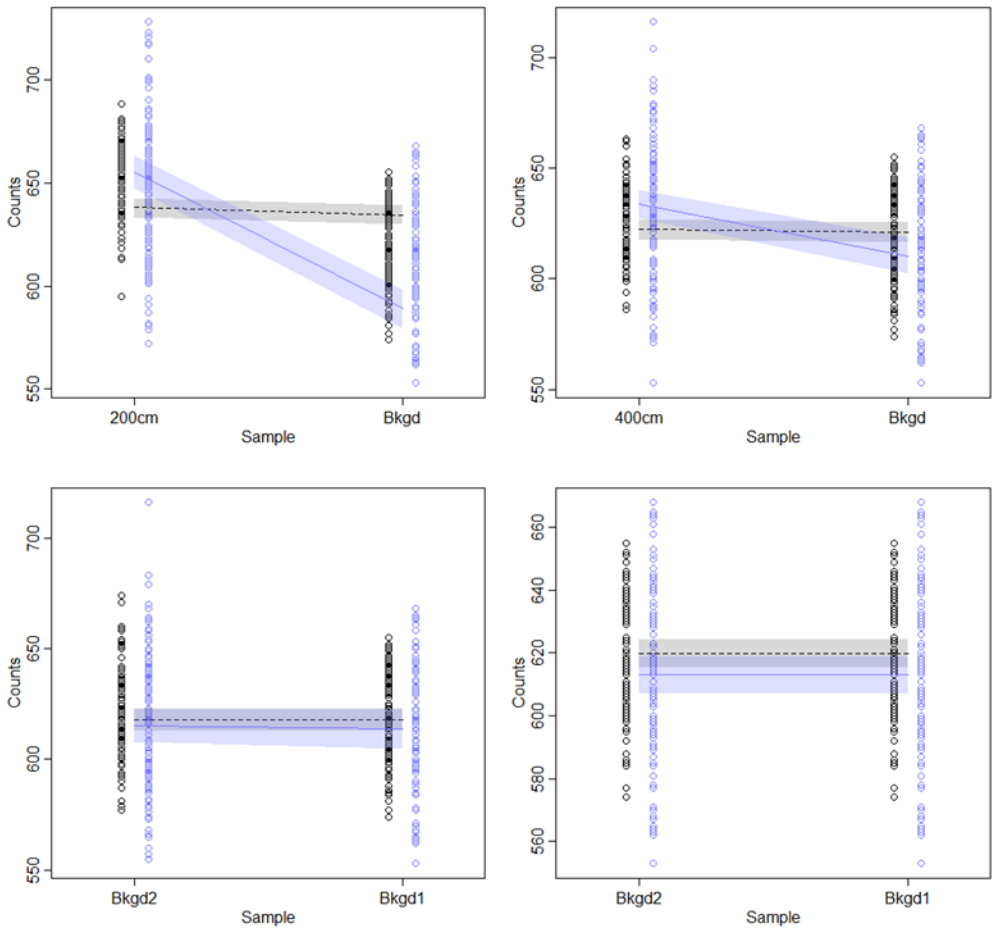

Figure B-17. Symmetry plots comparing Bkgd1 (labeled Bkgd/Bkgd1) to the 200cm source distance (top left) dataset, the 400cm source distance dataset (top right), the Bkgd2 dataset (Bkgd2 in bottom left), and the Self-test (bottom right)

Table B-7. MAP Estimates for Parameters from the Interaction Model

	Unknown Sample Dataset							
	200cm		400cm		Background 2		Self	
parameter θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ
α	637.19	3.33	621.61	3.3	618.05	3.39	620.29	3.21
β_{SD5}	0.32	0.13	0.25	0.13	-0.06	0.13	-0.17	0.15
β_{Bkgd_Sample}	-0.53	0.99	0.03	0.99	0.17	0.99	0.1	0.99
$\beta_{Bkgd\_Sample,SD5}$	-1.16	0.11	-0.47	0.1	-0.03	0.1	0	0.1
σ	25.98	0.87	24.45	0.82	24.98	0.83	23.21	0.77

Table B-8. Calculated γ for Given Source Types

Source Type	γ	
	No Source	Source
200cm	-0.84	0.32
400cm	-0.22	0.24
Background 2	-0.09	-0.06
Self	-0.17	-0.17

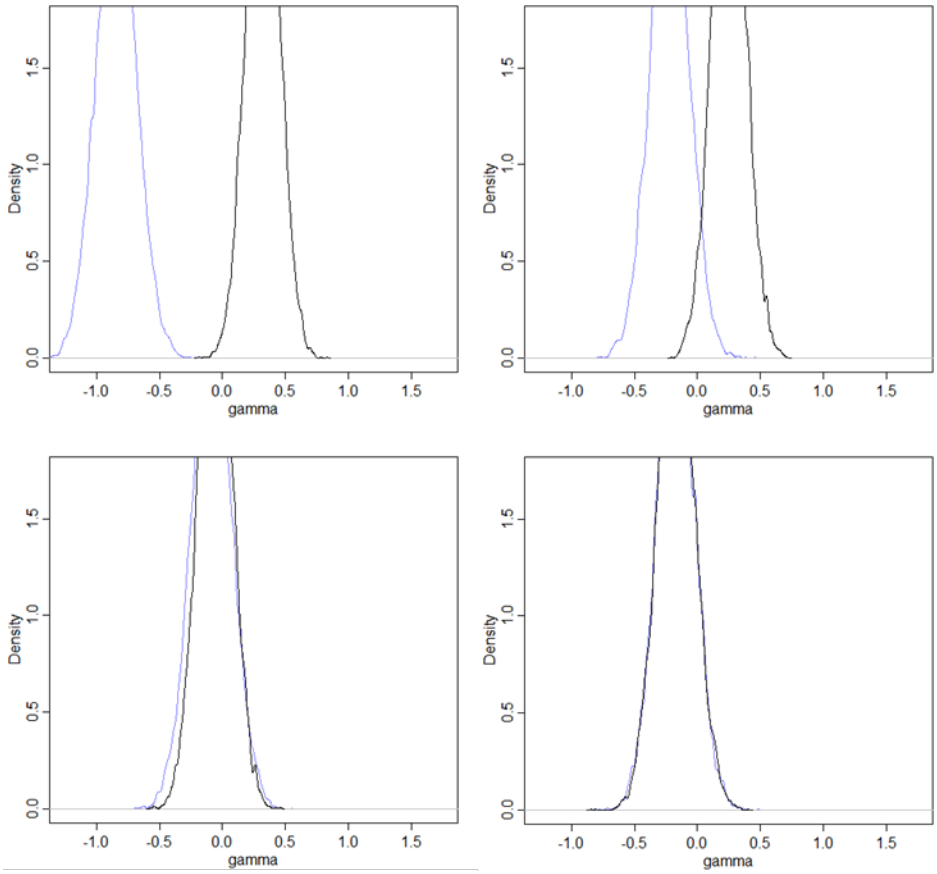

Figure B-18. Comparisons of marginal distributions for Bkgd1 (violet) versus source type 400cm (top left), 400cm (top right), Bkgd2 (bottom left), and Self (bottom right)

Table B-9. Means of the Marginal Distributions for a Given Source Type

Source Type	μ_γ	
	No Source	Source
200cm	-0.8435	0.3201
400cm	-0.2219	0.2474
Background 2	-0.0868	-0.0583
Self	-0.1744	-0.1713

$n = 113$ *Data*

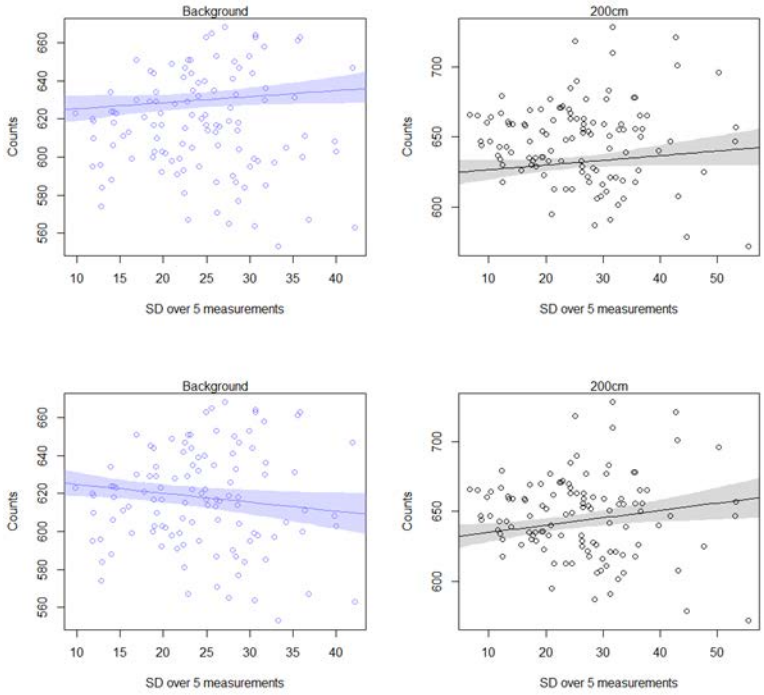

Figure B-19. Regression plots for the 200cm source distance

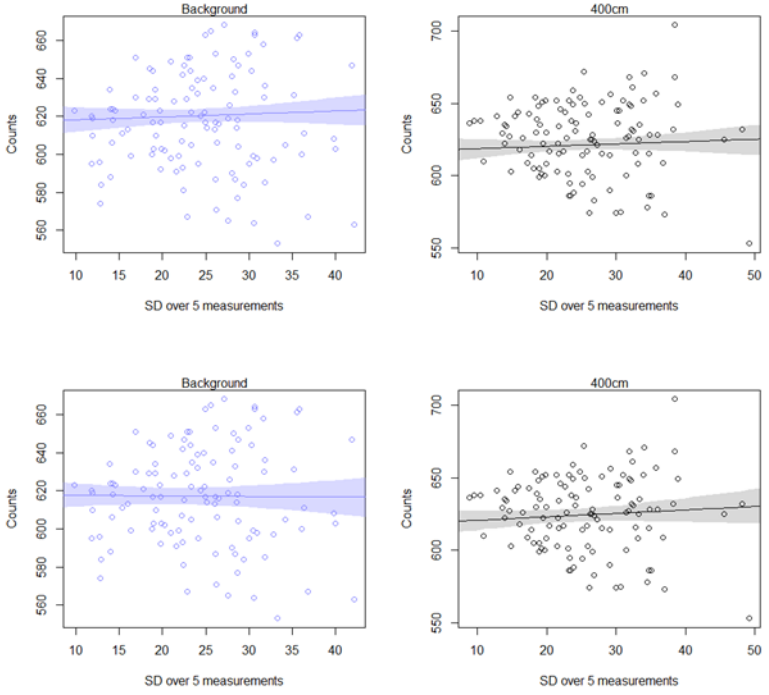

Figure B-20. Regression plots for the 400cm source distance

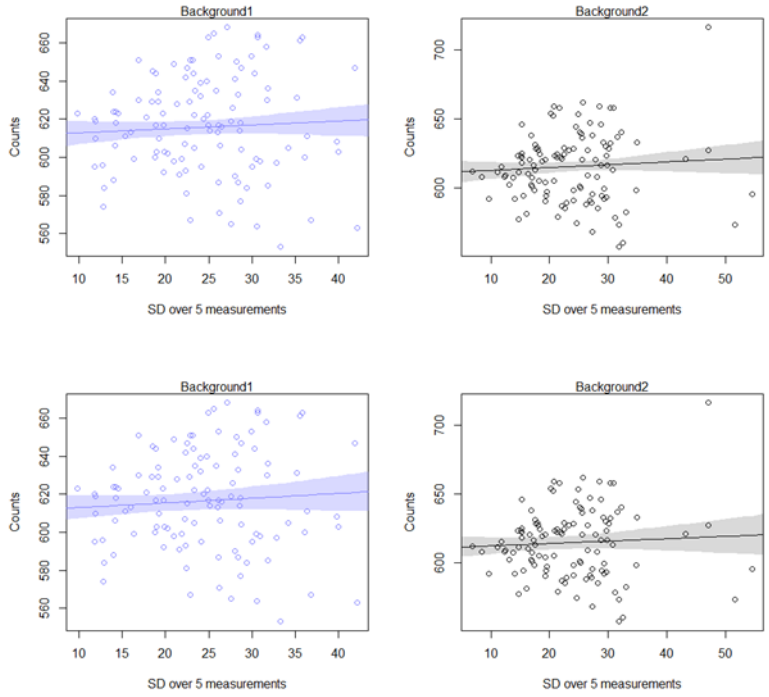

Figure B-21. Regression plots for the Bkgd2 source distance

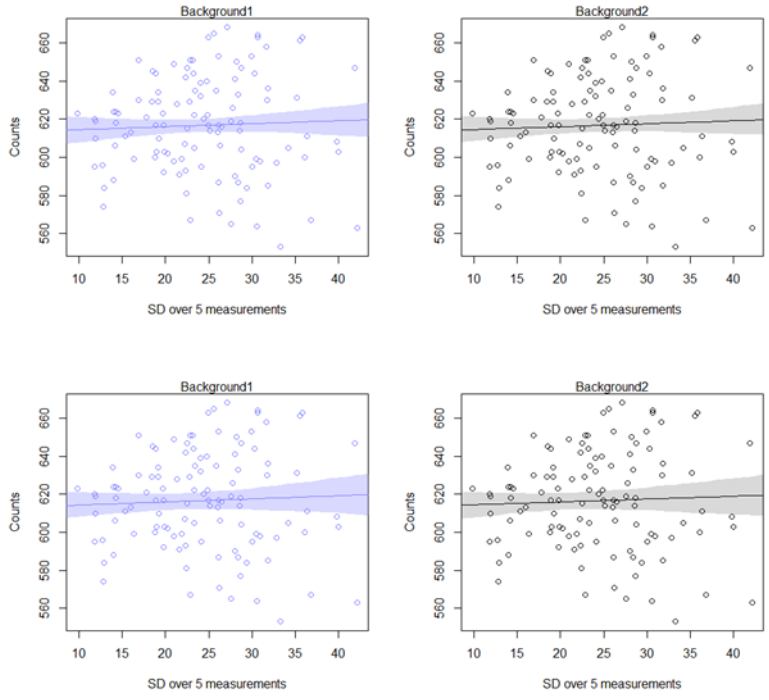

Figure B-22. Regression plots for the Self-test

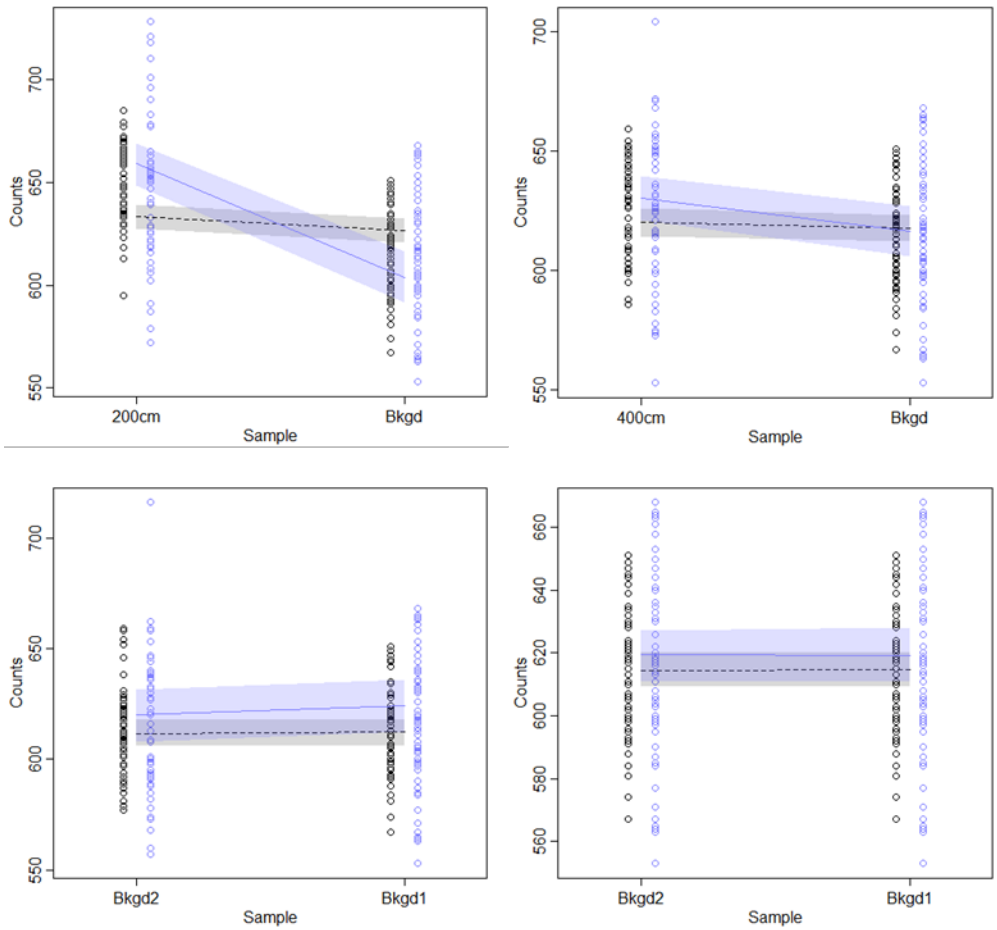

Figure B-23. Symmetry plots comparing Bkgd1 (labeled Bkgd/Bkgd1) to the 200cm source distance (top left) dataset, the 400cm source distance dataset (top right), the Bkgd2 dataset (Bkgd2 in bottom left), and the Self-test (bottom right)

Table B-10. MAP Estimates for Parameters from the Interaction Model

	Unknown Sample Dataset							
	200cm		400cm		Background 2		Self	
parameter θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ
α	629375	4.87	618.08	4.99	610.42	4.76	612.86	5.11
β_{SD5}	0.53	0.19	0.25	0.2	0.17	0.2	0.16	0.22
β_{Bkgd_Sample}	-0.18	1	-0.02	0.99	0.1	0.99	0.06	0.99
$\beta_{Bkgd\_Sample,SD5}$	-1	0.14	-0.28	0.14	0.07	0.14	0	0.14
σ	27.86	1.32	25.08	1.22	25.59	1.2	26.01	1.23

Table B-11. Calculated γ for Given Source Types

Source Type	γ	
	No Source	Source
200cm	-0.47	0.53
400cm	-0.04	0.25
Background 2	0.24	0.17
Self	0.16	0.16

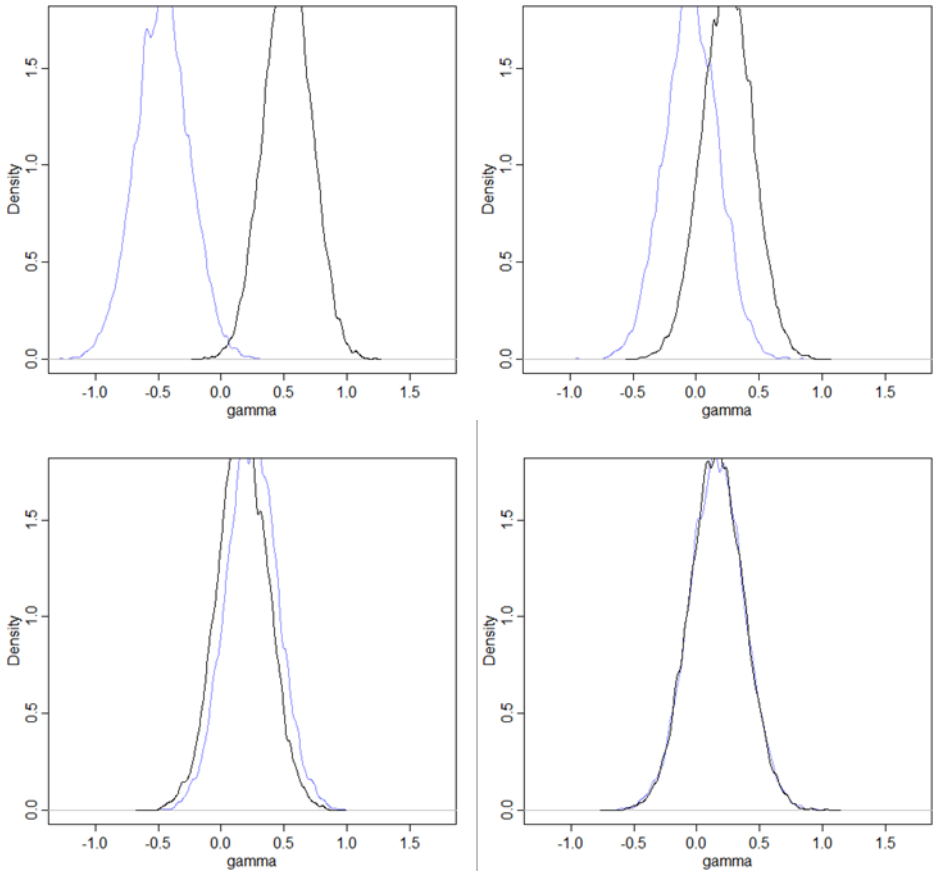

Figure B-24. Comparisons of marginal distributions for Bkgd1 (violet) versus source type 400cm (top left), 400cm (top right), Bkgd2 (bottom left), and Self (bottom right)

Table B-12. Means of the Marginal Distributions for a Given Source Type

Source Type	μ_γ	
	No Source	Source
200cm	-0.4673	0.5286
400cm	-0.0374	0.2464
Background 2	0.2434	0.171
Self	0.1582	0.1574

$n = 57$ *Data*

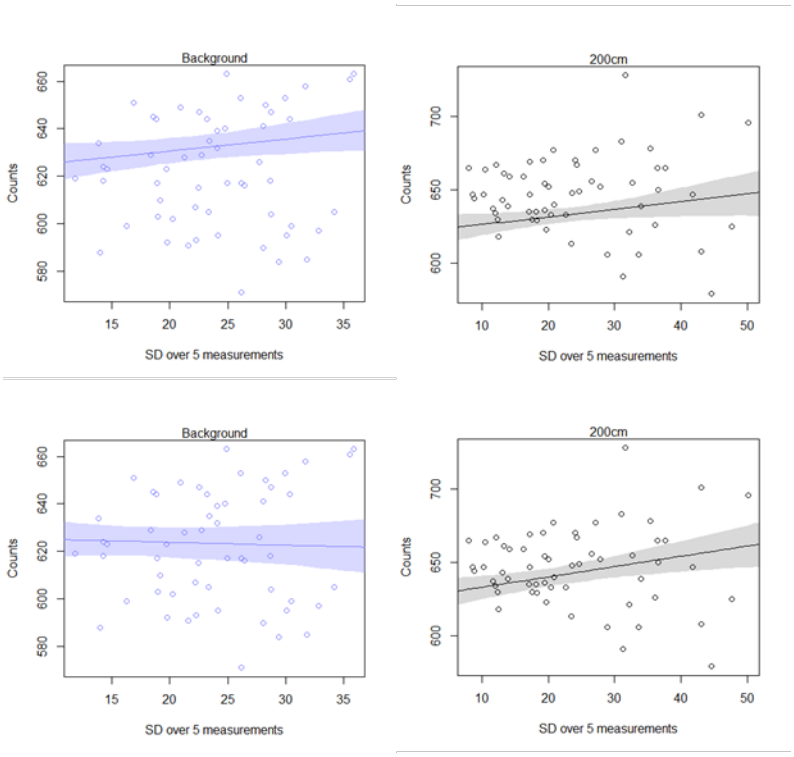

Figure B-25. Regression plots for the 200cm source distance

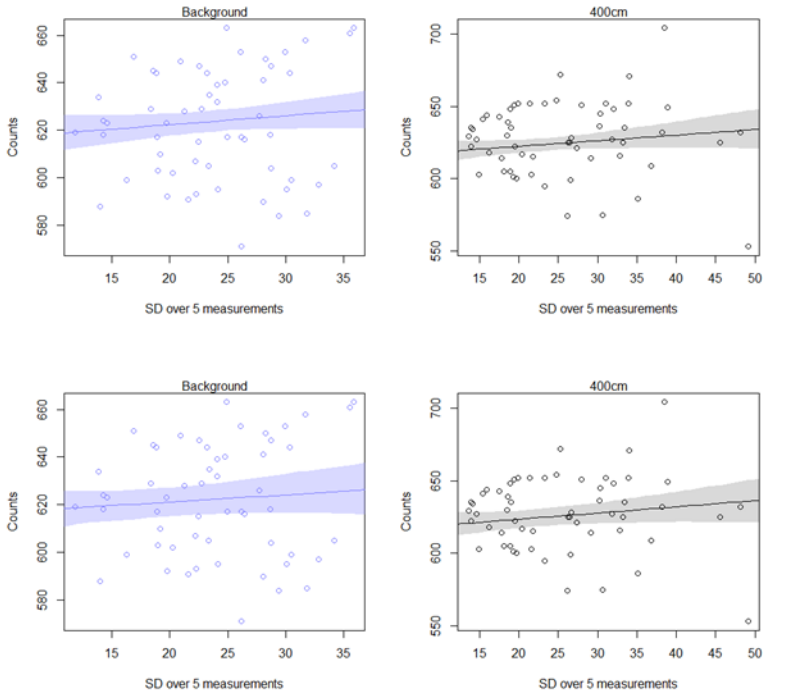

Figure B-26. Regression plots for the 400cm source distance

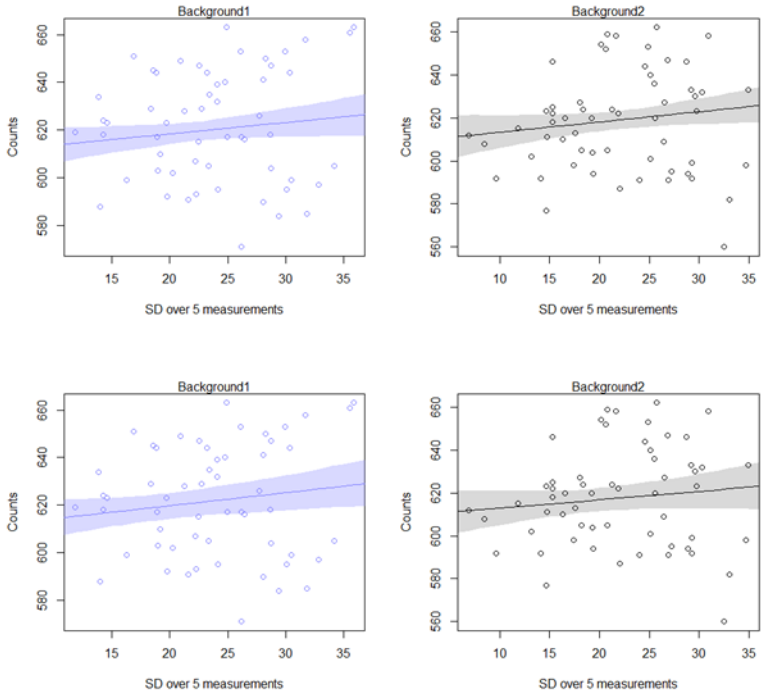

Figure B-27. Regression plots for the Bkgd2 source distance

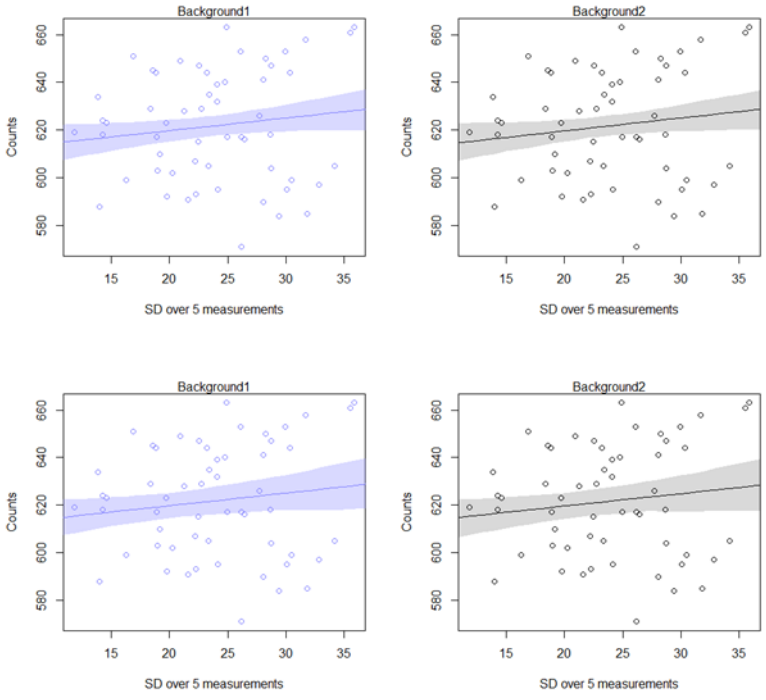

Figure B-28. Regression plots for the Self-test

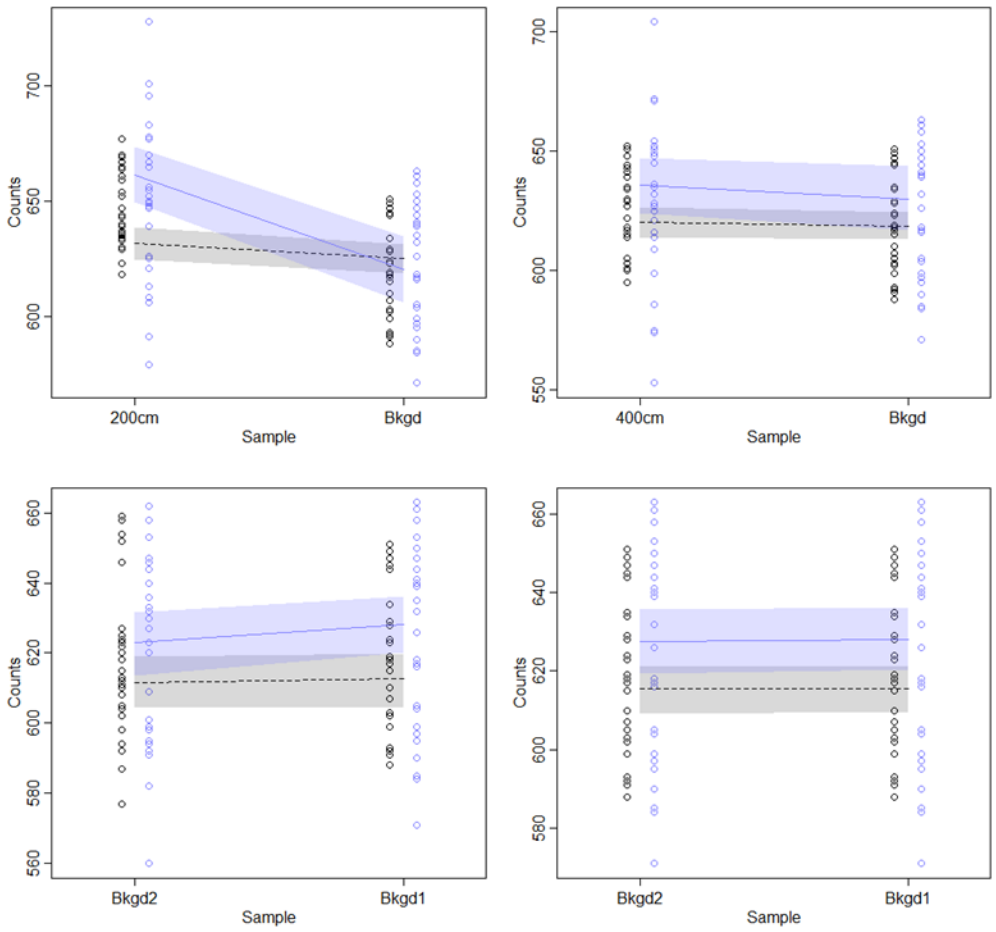

Figure B-29. Symmetry plots comparing Bkgd1 (labeled Bkgd/Bkgd1) to the 200cm source distance (top left) dataset, the 400cm source distance dataset (top right), the Bkgd2 dataset (Bkgd2 in bottom left), and the Self-test (bottom right)

Table B-13. MAP Estimates for Parameters from the Interaction Model

	Unknown Sample Dataset							
	200cm		400cm		Background 2		Self	
parameter θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ
α	626.21	5.77	615.15	6.15	608.72	6.1	609.17	6.49
β_{SD5}	0.69	0.23	0.42	0.24	0.39	0.28	0.52	0.28
β_{Bkgd_Sample}	-0.09	1	-0.01	1	0.04	1	0.03	1
$\beta_{Bkgd\_Sample,SD5}$	-0.82	0.19	-0.12	0.18	0.15	0.18	0.01	0.18
σ	25.58	1.72	24.62	1.64	23.23	1.54	23031	1.55

Table B-14. Calculated γ for Given Source Types

Source Type	γ	
	No Source	Source
200cm	-0.13	0.69
400cm	0.3	0.42
Background 2	0.54	0.39
Self	0.53	0.52

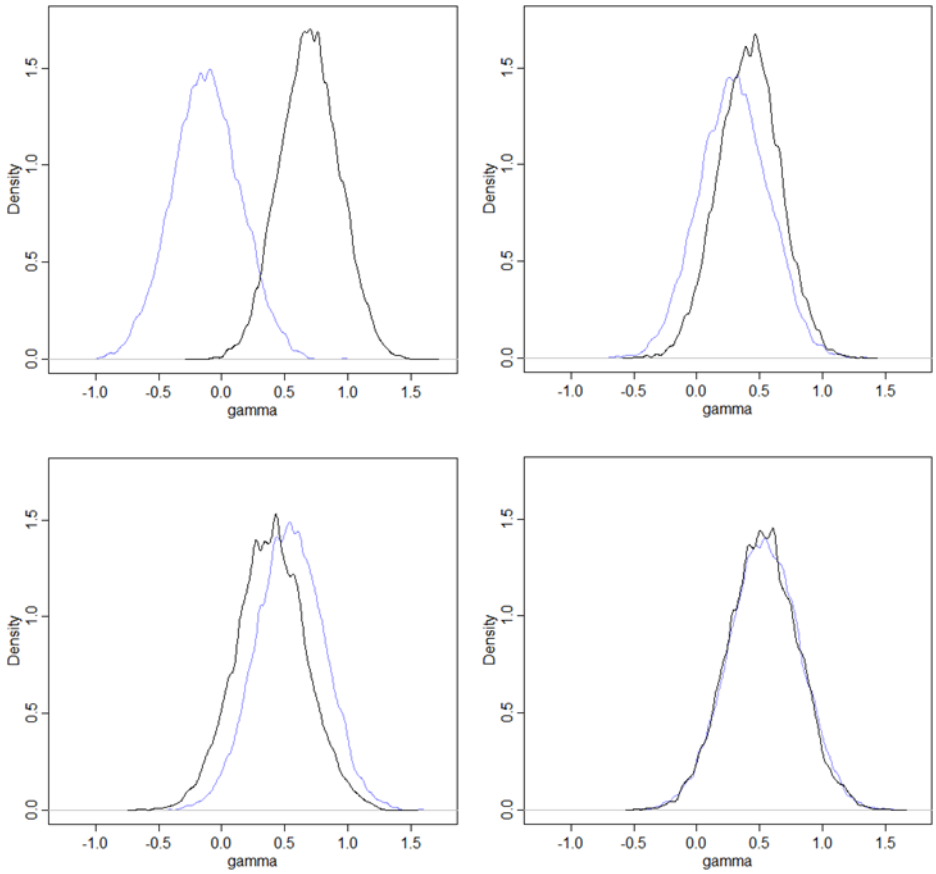

Figure B-30. Comparisons of marginal distributions for Bkgd1 (violet) versus source type 400cm (top left), 400cm (top right), Bkgd2 (bottom left), and Self (bottom right)

Table B-15. Means of the Marginal Distributions for a Given Source Type

Source Type	μ_γ	
	No Source	Source
200cm	-0.1235	0.6956
400cm	0.2953	0.4144
Background 2	0.542	0.3907
Self	0.5353	0.5253

$n = 29$ *Data*

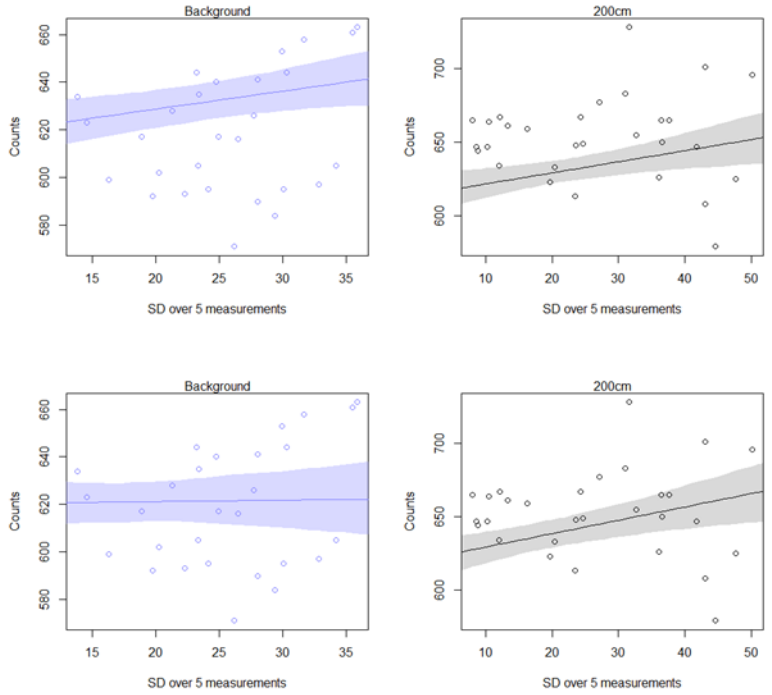

Figure B-31. Regression plots for the 200cm source distance

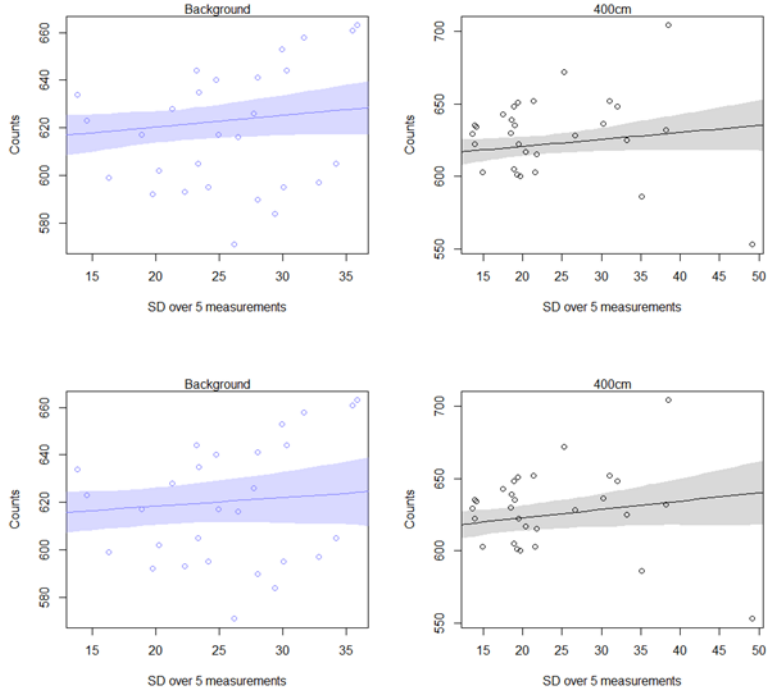

Figure B-32. Regression plots for the 400cm source distance

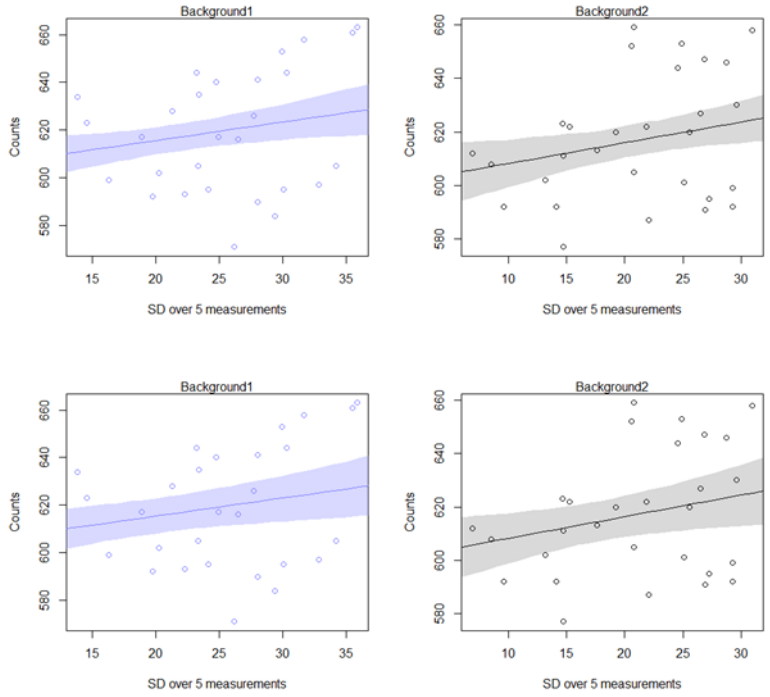

Figure B-33. Regression plots for the Bkgd2 source distance

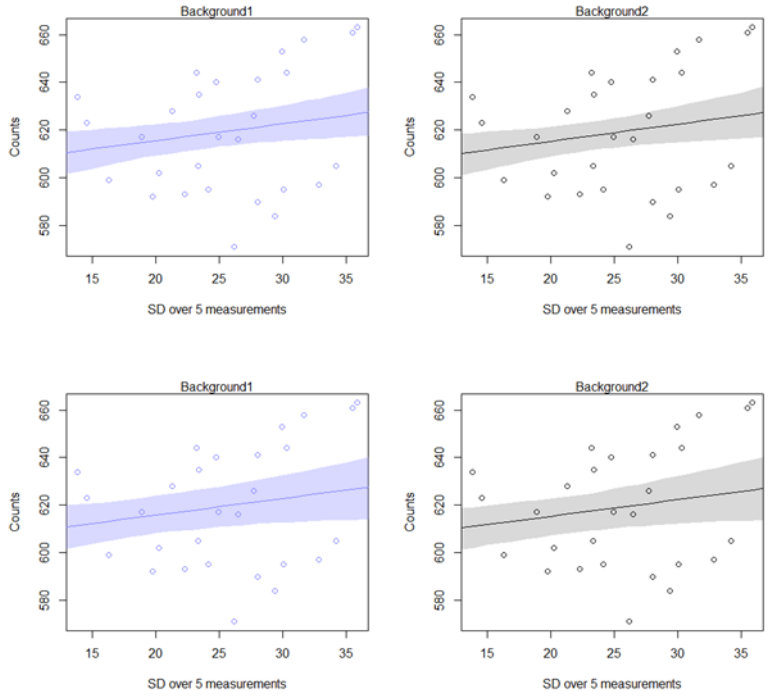

Figure B-34. Regression plots for the Self-test

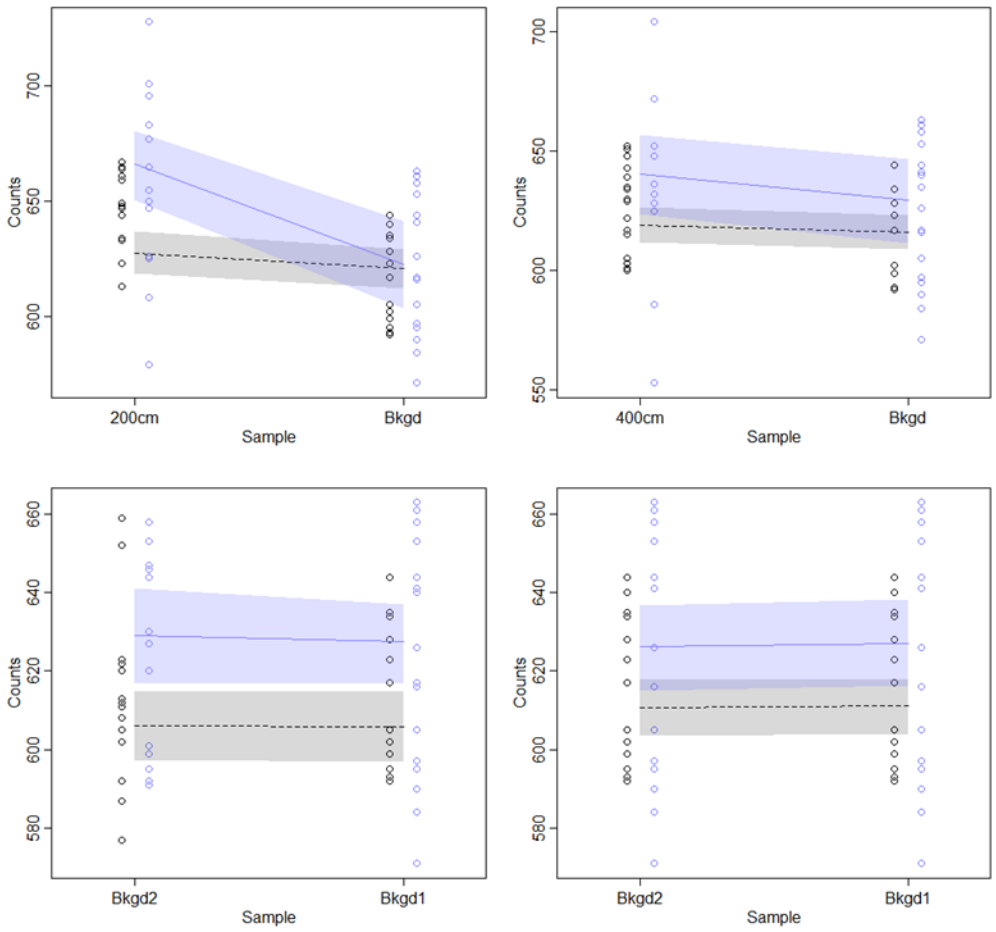

Figure B-35. Symmetry plots comparing Bkgd1 (labeled Bkgd/Bkgd1) to the 200cm source distance (top left) dataset, the 400cm source distance dataset (top right), the Bkgd2 dataset (Bkgd2 in bottom left), and the Self-test (bottom right)

Table B-16. MAP Estimates for Parameters from the Interaction Model

	Unknown Sample Dataset							
	200cm		400cm		Background 2		Self	
parameter θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ
α	620.04	7.39	611.01	7.4	600.23	7.09	601.41	7.71
β_{SD5}	0.92	0.28	0.58	0.33	0.8	0.35	0.7	0.33
β_{Bkgd_Sample}	-0.07	1	-0.04	1	0	1	-0.01	1
$\beta_{Bkgd\_Sample,SD5}$	-0.86	0.27	-0.21	0.26	-0.04	0.25	0.02	0.24
σ	29.08	2079	26.61	2.49	23.19	2.16	24.37	2.26

Table B-17. Calculated γ for Given Source Types

Source Type	γ	
	No Source	Source
200cm	0.06	0.92
400cm	0.37	0.58
Background 2	0.76	0.8
Self	0.72	0.7

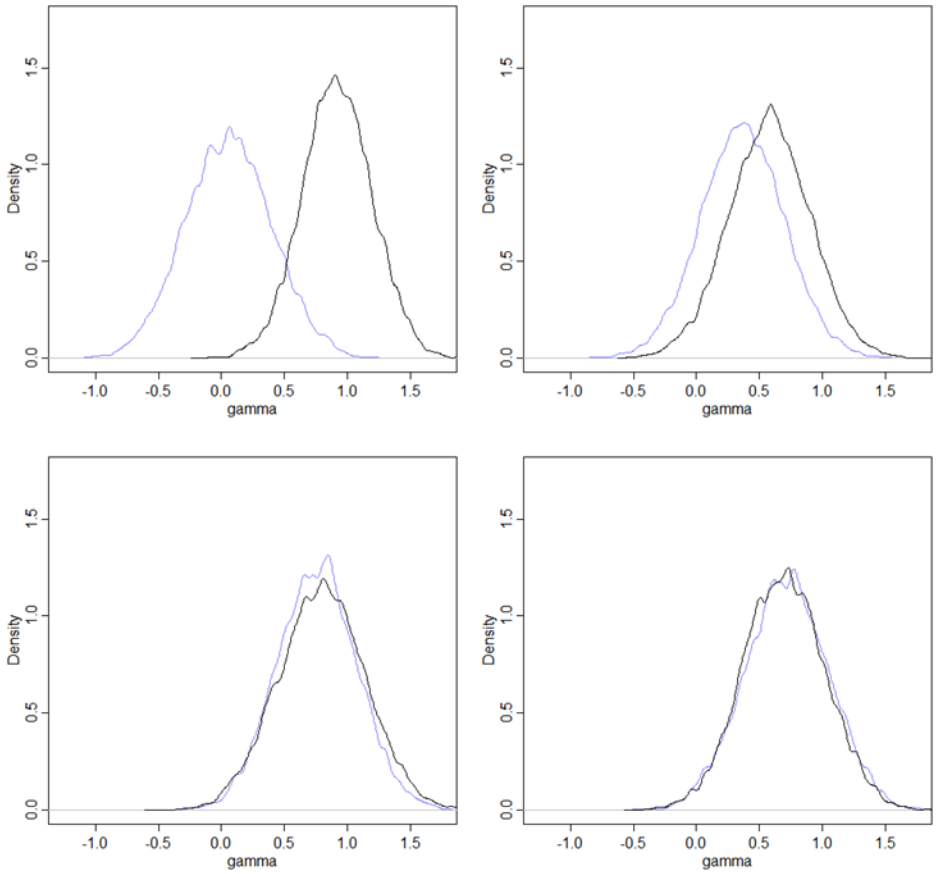

Figure B-36. Comparisons of marginal distributions for Bkgd1 (violet) versus source type 400cm (top left), 400cm (top right), Bkgd2 (bottom left), and Self (bottom right)

Table B-18. Means of the Marginal Distributions for a Given Source Type

Source Type	μ_γ	
	No Source	Source
200cm	0.0561	0.9187
400cm	0.3659	0.579
Background 2	0.76	0.7966
Self	0.716	0.6958

$n = 15$ *Data*

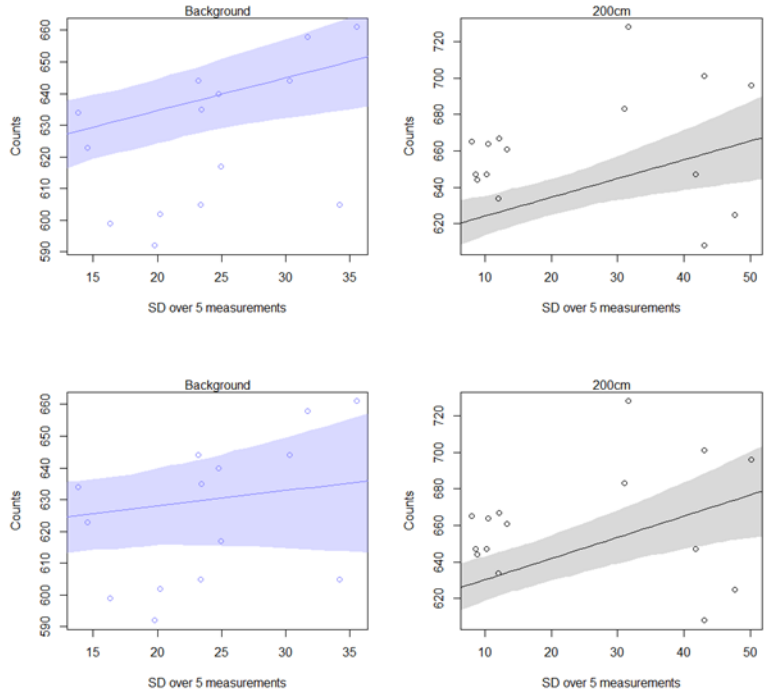

Figure B-37. Regression plots for the 200cm source distance

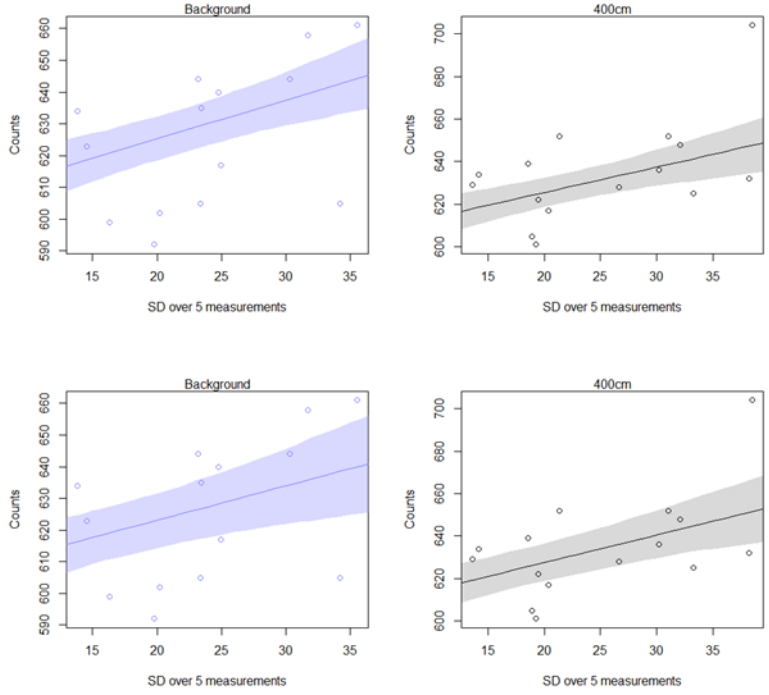

Figure B-38. Regression plots for the 400cm source distance

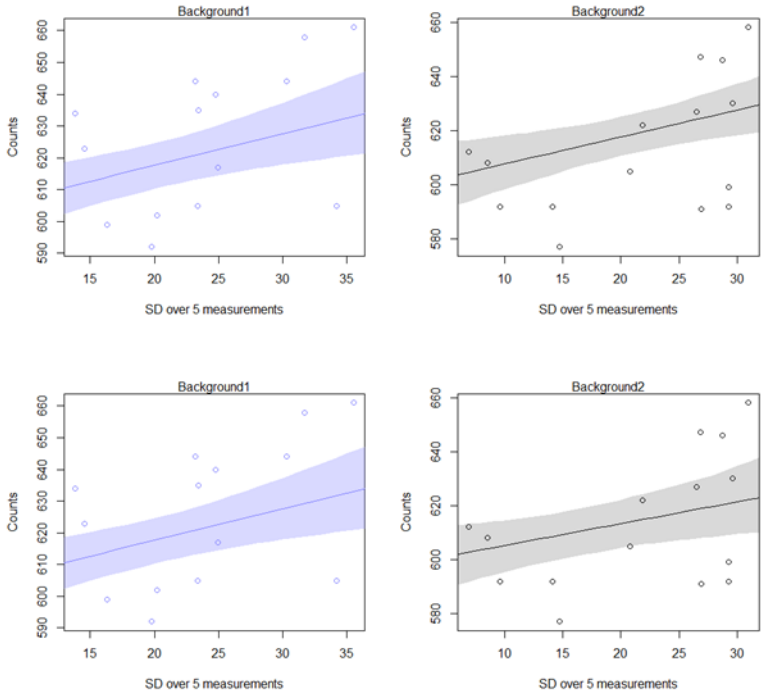

Figure B-39. Regression plots for the Bkgd2 source distance

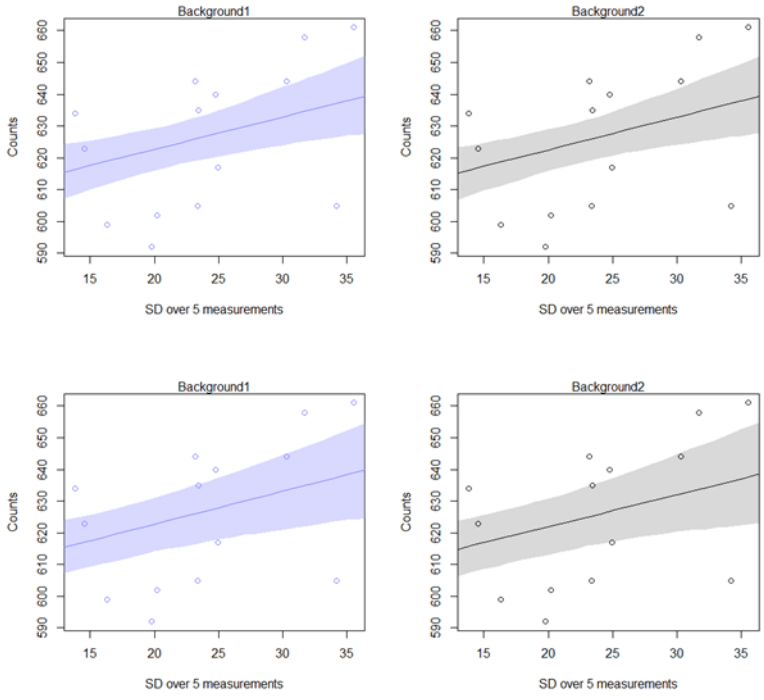

Figure B-40. Regression plots for the Self-test

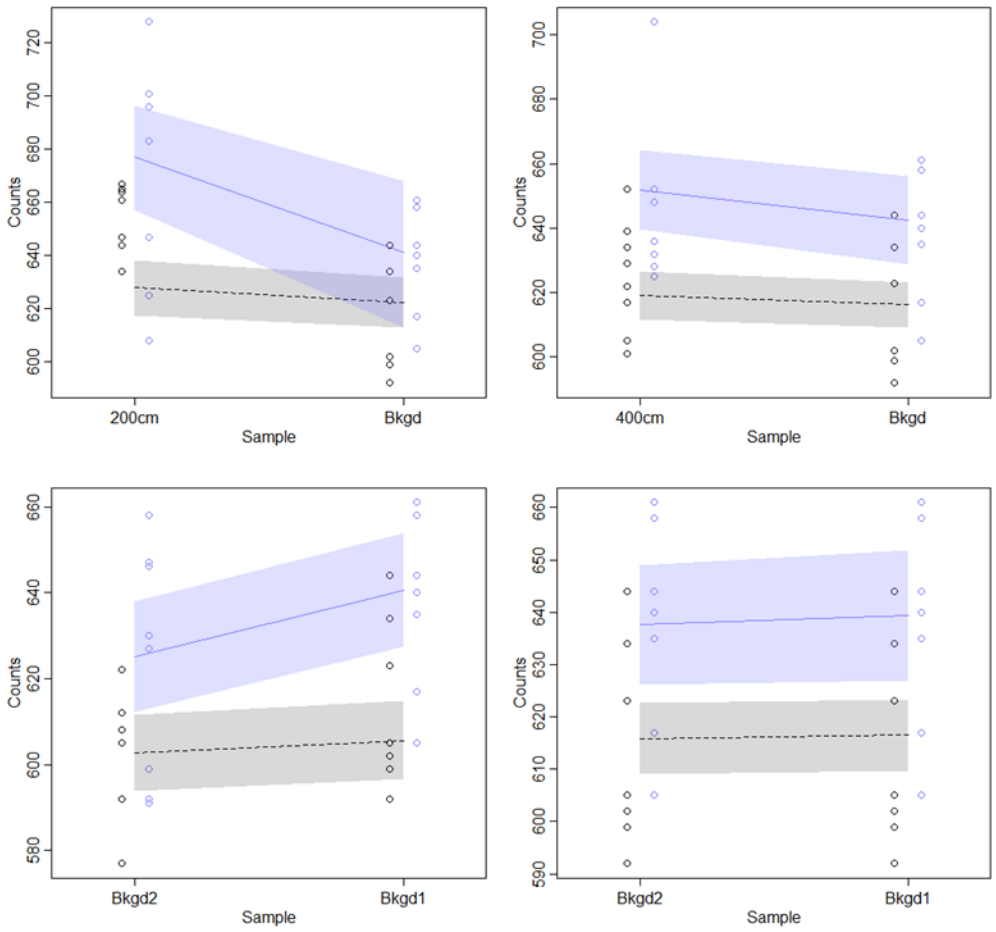

Figure B-41. Symmetry plots comparing Bkgd1 (labeled Bkgd/Bkgd1) to the 200cm source distance (top left) dataset, the 400cm source distance dataset (top right), the Bkgd2 dataset (Bkgd2 in bottom left), and the Self-test (bottom right)

Table B-19. MAP Estimates for Parameters from the Interaction Model

	Unknown Sample Dataset							
	200cm		400cm		Background 2		Self	
parameter θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ
α	618.53	7.95	601.55	7.47	597.32	7.27	601.82	7.49
β_{SD5}	1.17	0.32	1.3	0.32	0.79	0.35	1.02	0.35
β_{Bkgd_Sample}	-0.06	1	-0.02	1	0.02	1	-0.01	1
$\beta_{Bkgd\_Sample,SD5}$	-0.7	0.39	-0.22	0.28	0.42	0.3	0.04	0.28
σ	29.14	4	19.74	2.56	20.33	2.64	19.56	2.53

Table B-20. Calculated γ for Given Source Types

Source Type	γ	
	No Source	Source
200cm	0.47	1.17
400cm	1.08	1.3
Background 2	1.21	0.79
Self	0.72	0.7

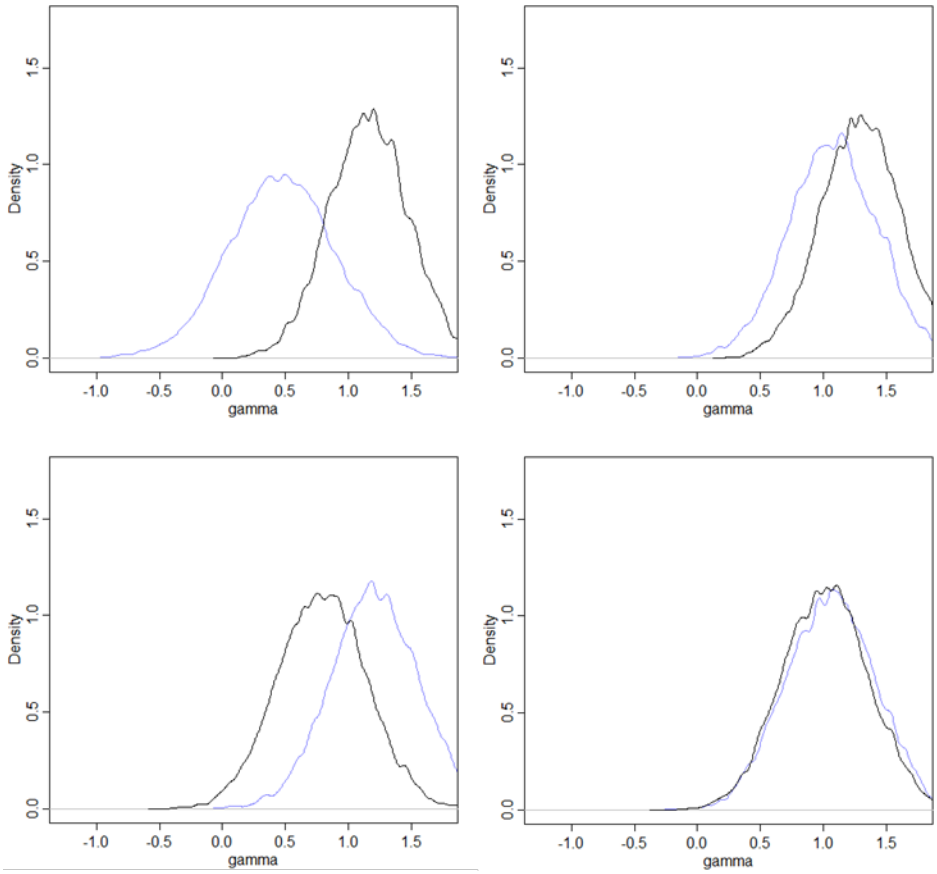

Figure B-42. Comparisons of marginal distributions for Bkgd1 (violet) versus source type 400cm (top left), 400cm (top right), Bkgd2 (bottom left), and Self (bottom right)

Table B-21. Means of the Marginal Distributions for a Given Source Type

Source Type	μ_γ	
	No Source	Source
200cm	0.469	1.161
400cm	1.076	1.299
Background 2	1.21	0.7966
Self	1.056	1.017

$n = 8$ *Data*

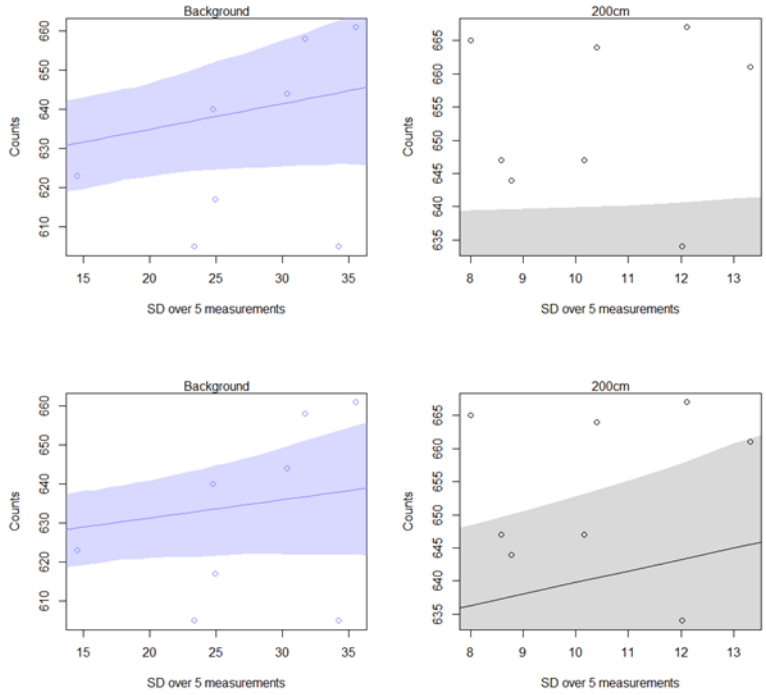

Figure B-43. Regression plots for the 200cm source distance

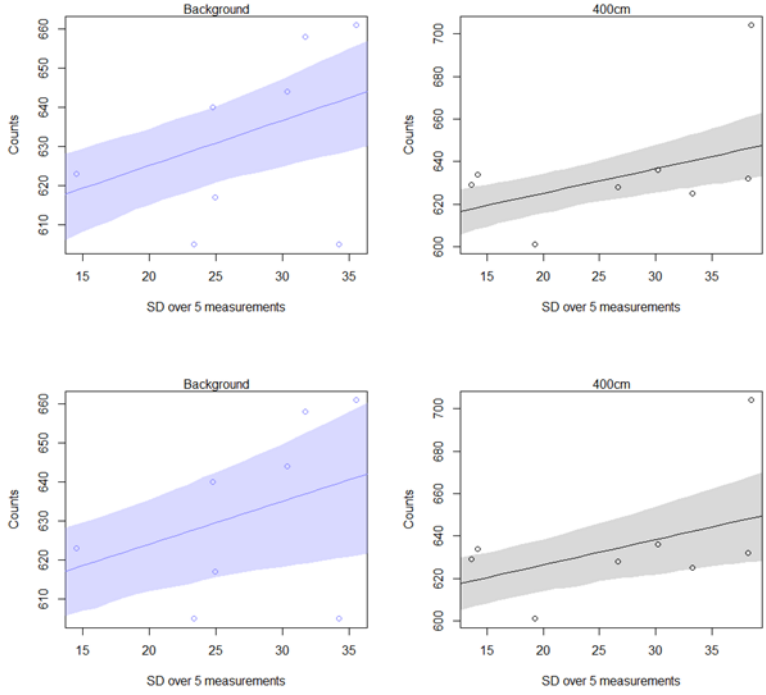

Figure B-44. Regression plots for the 400cm source distance

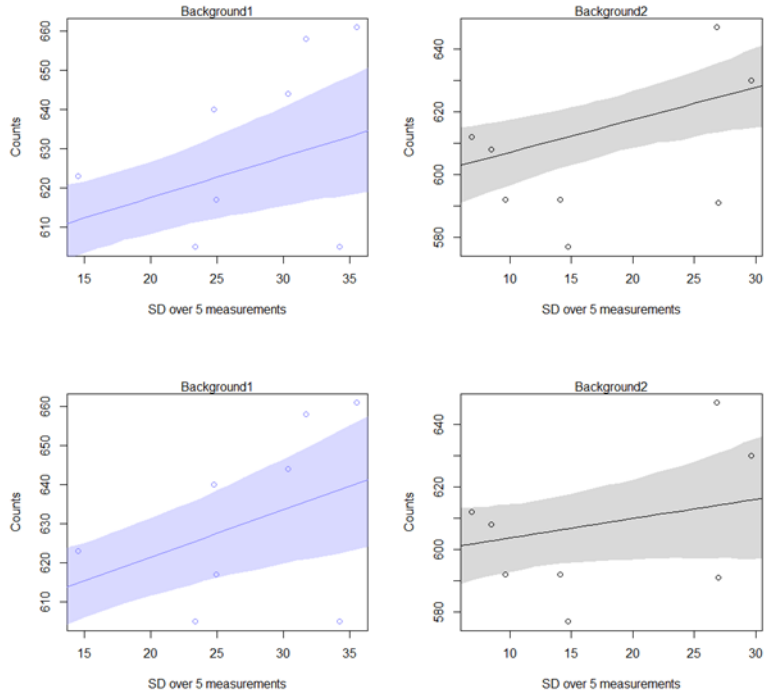

Figure B-45. Regression plots for the Bkgd2 source distance

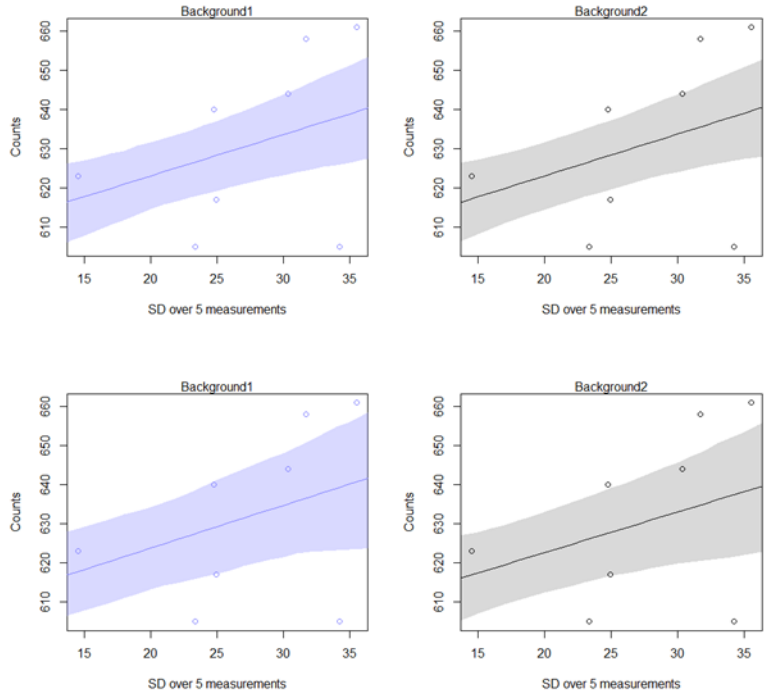

Figure B-46. Regression plots for the Self-test

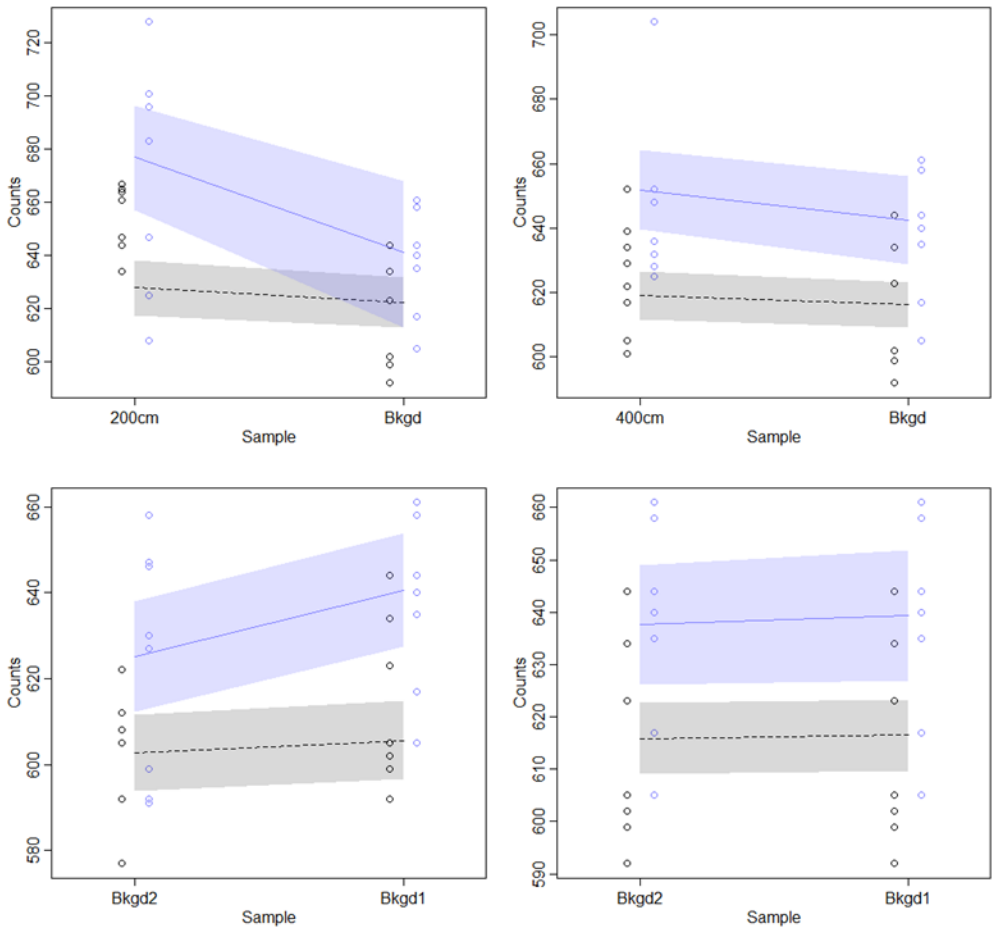

Figure B-47. Symmetry plots comparing Bkgd1 (labeled Bkgd/Bkgd1) to the 200cm source distance (top left) dataset, the 400cm source distance dataset (top right), the Bkgd2 dataset (Bkgd2 in bottom left), and the Self-test (bottom right)

Table B-22. MAP Estimates for Parameters from the Interaction Model

	Unknown Sample Dataset							
	200cm		400cm		Background 2		Self	
parameter θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ
α	622.21	7.56	602.25	8.52	597.4	7.66	602.11	8.64
β_{SD5}	1.74	0.66	1.2	0.37	0.63	0.45	1.03	0.36
β_{Bkgd_Sample}	-0.08	1	-0.01	1	0.02	1	-0.01	1
$\beta_{Bkgd\_Sample,SD5}$	-1.27	0.59	-0.1	0.36	0.58	0.39	0.05	0.32
σ	18.89	4.01	21.71	3.87	19.49	3.48	19.17	3.4

Table B-23. Calculated γ for Given Source Types

Source Type	γ	
	No Source	Source
200cm	0.47	1.17
400cm	1.1	1.2
Background 2	1.21	0.63
Self	1.08	1.03

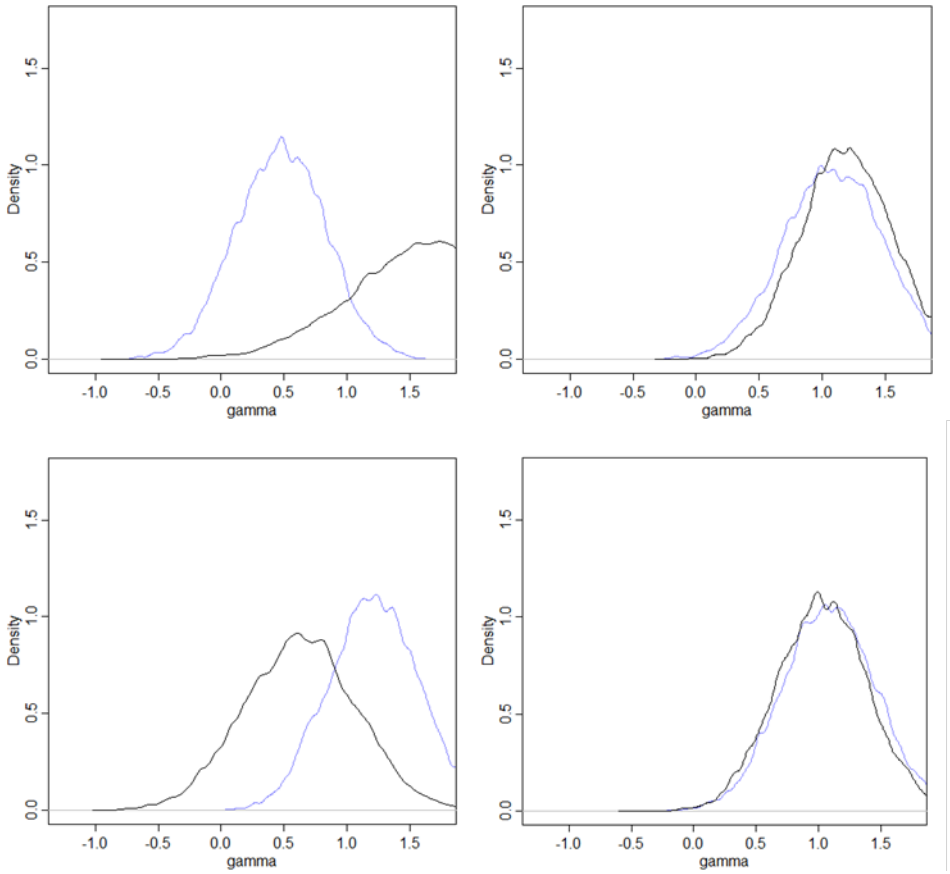

Figure B-48. Comparisons of marginal distributions for Bkgd1 (violet) versus source type 400cm (top left), 400cm (top right), Bkgd2 (bottom left), and Self (bottom right)

Table B-24. Means of the Marginal Distributions for a Given Source Type

Source Type	μ_γ	
	No Source	Source
200cm	0.4716	1.745
400cm	1.091	1.195
Background 2	1.213	0.6275
Self	1.085	1.031

$n = 4$ *Data*

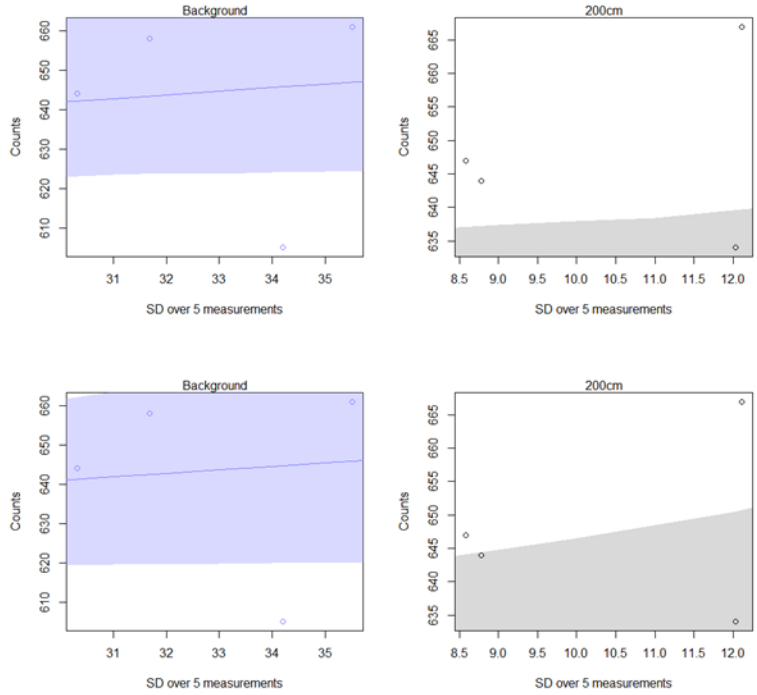

Figure B-49. Regression plots for the 200cm source distance

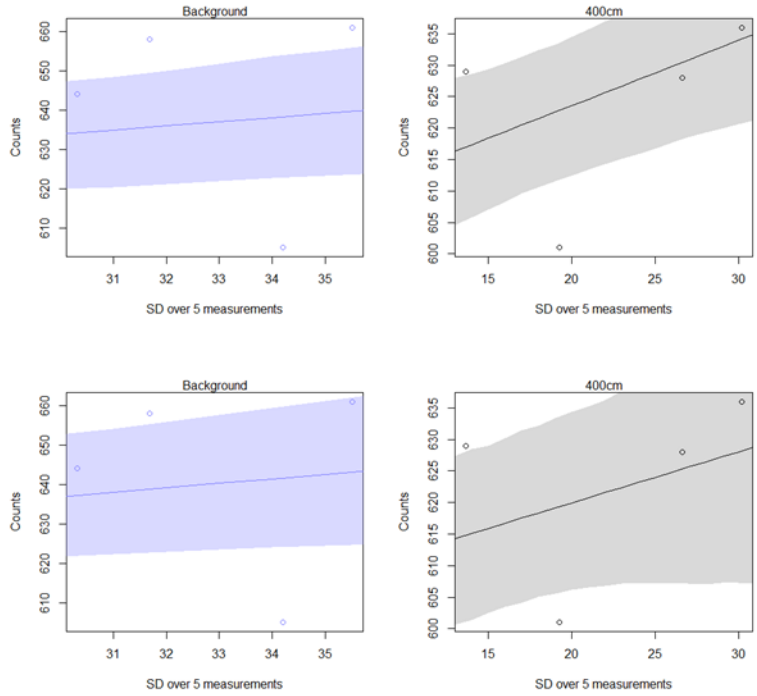

Figure B-50. Regression plots for the 400cm source distance

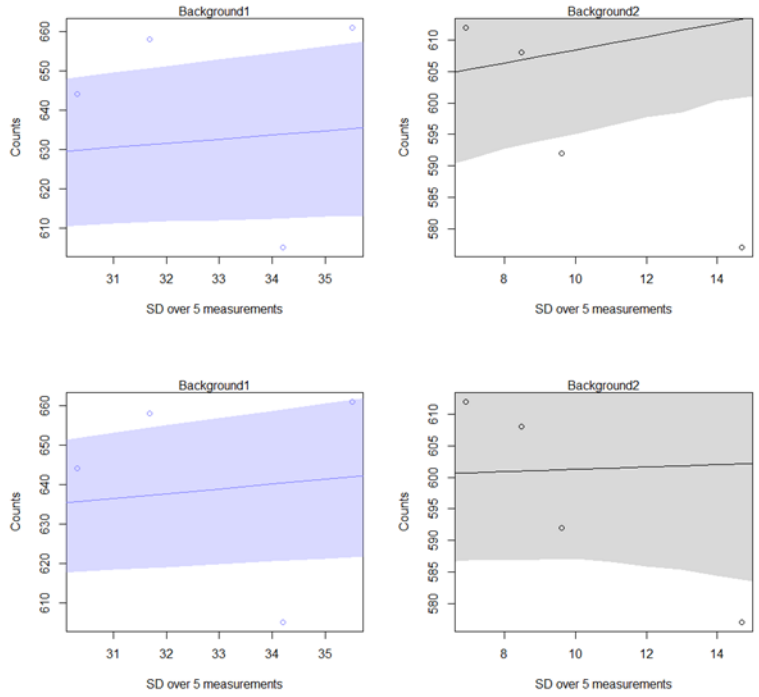

Figure B-51. Regression plots for the Bkgd2 source distance

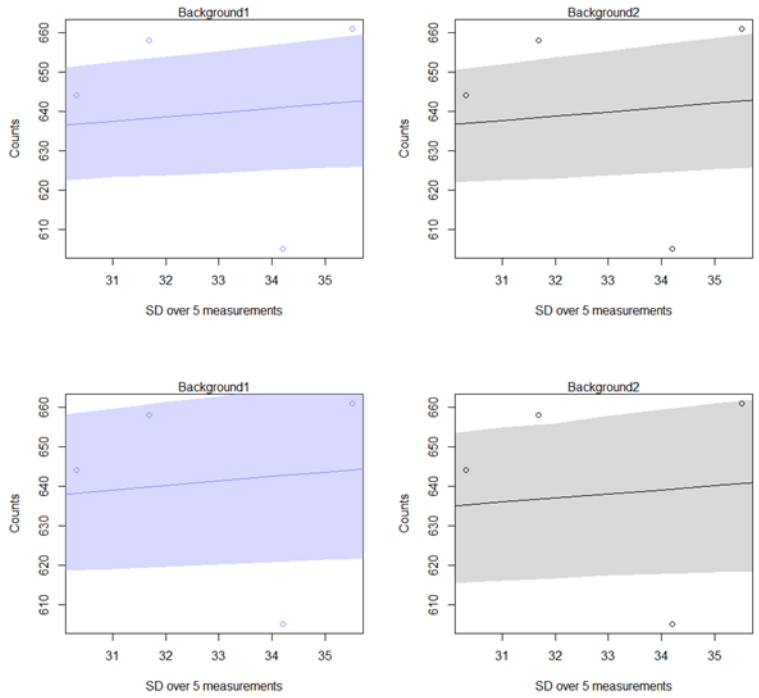

Figure B-52. Regression plots for the Self-test

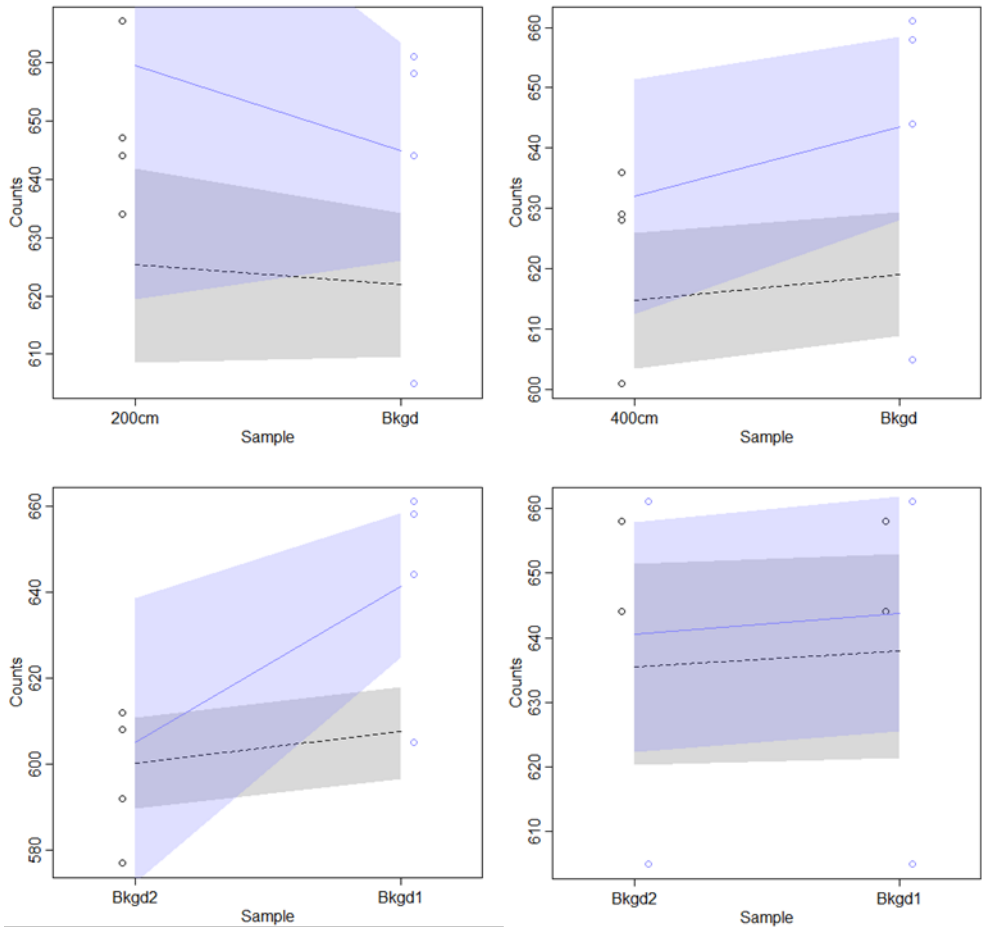

Figure B-53. Symmetry plots comparing Bkgd1 (labeled Bkgd/Bkgd1) to the 200cm source distance (top left) dataset, the 400cm source distance dataset (top right), the Bkgd2 dataset (Bkgd2 in bottom left), and the Self-test (bottom right)

Table B-25. MAP Estimates for Parameters from the Interaction Model

	Unknown Sample Dataset							
	200cm		400cm		Background 2		Self	
parameter θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ	μ_θ	σ_θ
α	614.46	9.38	603.53	8.97	599.33	7.98	603.78	9.59
β_{SD5}	1.24	0.72	0.81	0.47	0.18	0.68	1.05	0.41
β_{Bkgd_Sample}	-0.01	1	0.01	1	0.03	1	-0.01	1
$\beta_{Bkgd\_Sample,SD5}$	-0.36	0.71	0.32	0.42	1.02	0.63	0.09	0.43
σ	23.23	7.42	18.29	4.6	19.37	5.02	22.83	5.76

Table B-26. Calculated γ for Given Source Types

Source Type	γ	
	No Source	Source
200cm	0.88	1.24
400cm	1.13	0.81
Background 2	1.2	0.18
Self	1.14	1.05

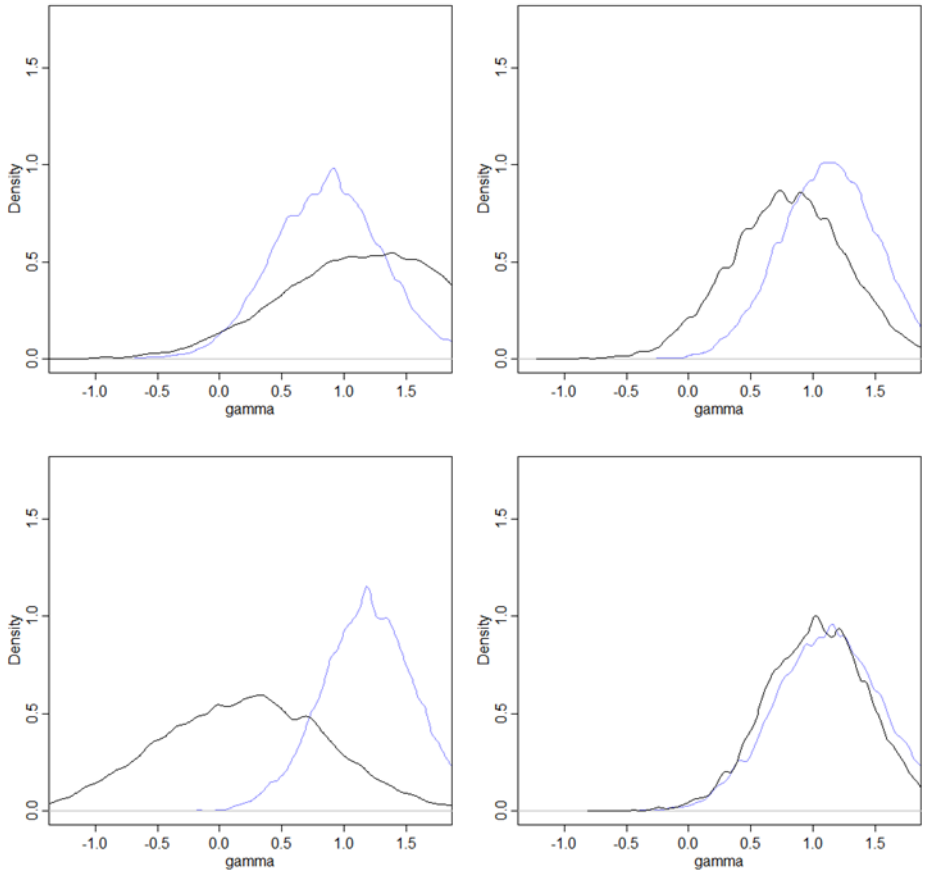

Figure B-54. Comparisons of marginal distributions for Bkgd1 (violet) versus source type 400cm (top left), 400cm (top right), Bkgd2 (bottom left), and Self (bottom right)

Table B-27. Means of the Marginal Distributions for a Given Source Type

Source Type	μ_γ	
	No Source	Source
200cm	0.8709	1.239
400cm	1.129	0.8022
Background 2	1.199	0.1868
Self	1.126	1.038

APPENDIX C

*Detection Rates for Various Models for Given Pr**

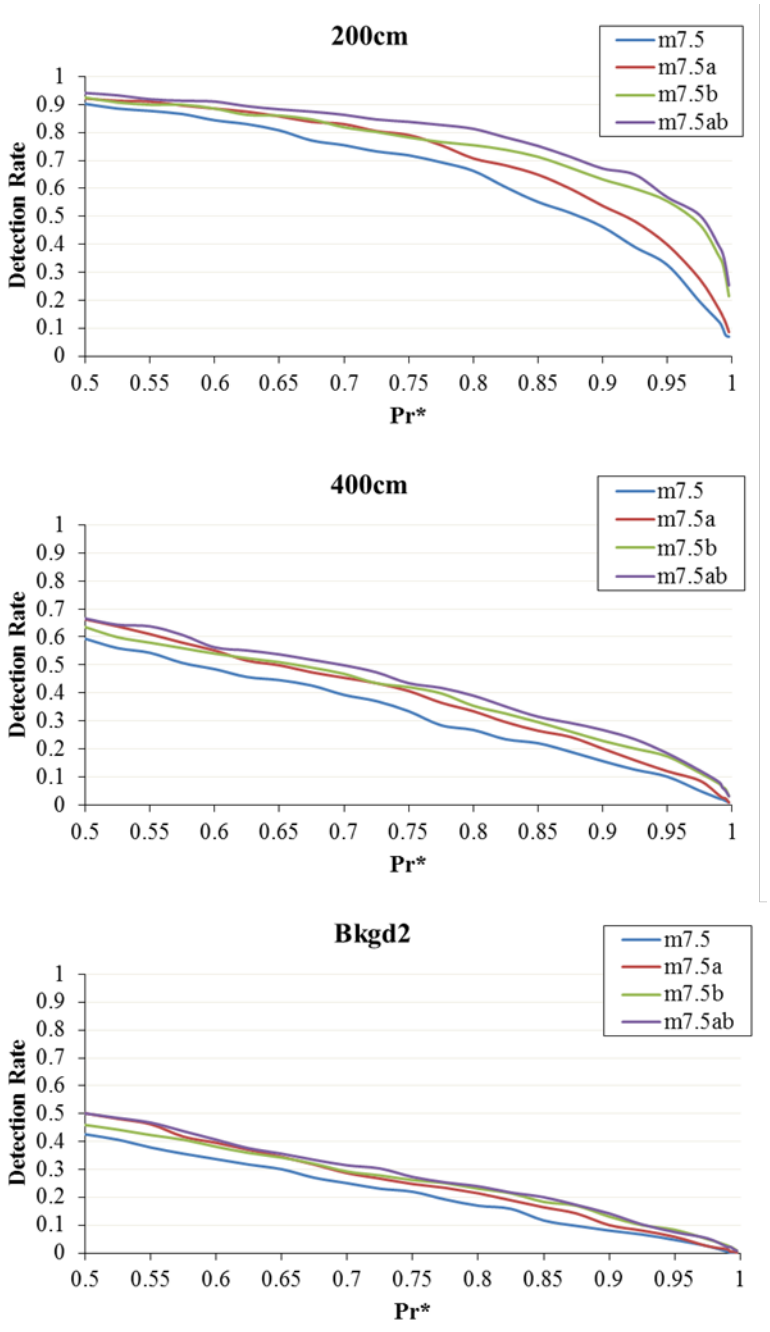

Figure C-1. Detection rates for various models on the given dataset

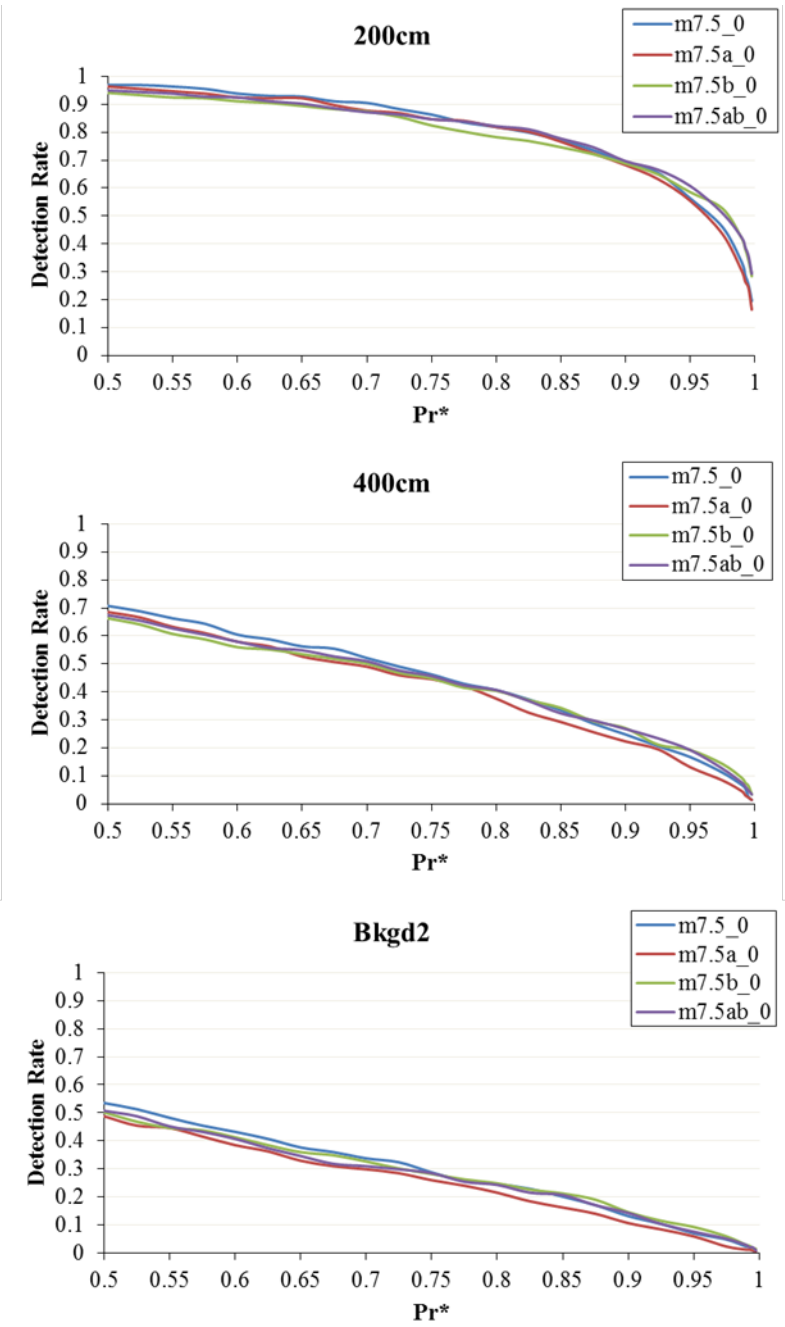

Figure C-1. Detection rates for various flipped predictor models on the given dataset