Detection of Sources of Harmful Radiation using Portable Sensors

Jay Jin

CMU-CS-16-115
May 2016

School of Computer Science
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA

Thesis Committee
Artur Dubrawski, Chair
Srinivas Narasimhan

Submitted in partial fulfillment of the requirements for the Degree of Master of Science

This research has been partially supported by the U.S. Department of Homeland Security, Domestic Nuclear Detection Office, under competitively awarded grant 2014-DN-077-ARI087-01, by the U.S. Department of Defense, Defense Threat Reduction Agency under award HDTRA1-13-1-0026, and by the National Science Foundation under award 1320347. This support does not constitute an express or implied endorsement on the part of the U.S. Government. Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the U.S. Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344.

Keywords: Nuclear Threat Detection, G-P detection method, List Mode Regression

Abstract

Detection and characterization of potentially harmful radiation sources is one of common problems encountered in the context of homeland security as well as in safeguarding of radioactive isotopes used in industrial or medical practices. Any improper or insecure storage of radioactive material may cause a substantial harm to humans and to the environment. The task of detecting relatively weak, potentially shielded sources, is particularly difficult when facing variability of patterns of background radiation. This is a common circumstance in cluttered urban environments. It is also in these environments where any improperly stored radioactive material may inflict the most extensive harm. In this thesis, we explore two algorithms that are useful in different situations. The first, the G-P detection method, works well with large sensors that can fit in vehicles. Its novelty is in making Poisson assumptions about the photon counts observed in gamma-ray spectra, while making Gaussian assumptions about their rates. The second, List Mode Regression, is tailored to small portable sensors which produce spectra with low photon counts. Both methods outperform current state of the art algorithms in their respective usage scenarios.

Acknowledgements

I would like to thank my advisor, Artur Dubrawski, for guiding me on various research challenges over the past three years. Artur has always pushed me to do my best work, and has been patient through both the highs and lows of research. He has always been there for me, whether it be a midnight paper review or practice presentation. Thank you for an incredible academic experience.

A huge thank you to Peter Huggins and Kyle Miller for their invaluable comments and suggestions. Without our weekly meetings, this research would not be possible. Thank you to fellow Auton Lab members Dougal Sutherland, Rob MacLachlan, and Prateek Tandon for their collaboration. I would also like to thank Simon Labov and Karl Nelson of Lawrence Livermore National Laboratory for their nuclear physics expertise and contributions.

Finally, I would like to thank family, friends, teammates, and coaches for all their support during my time here at CMU.

Contents

List of Figures

Introduction

The ability to detect nuclear materials is key in preventing the proliferation of nuclear weapons and nuclear terror attacks. Terrorist organizations and even certain countries acquiring nuclear weapons is a dangerous situation for the entire international community[12]. The smuggling of weapons such as tactical nuclear weapons is a very real threat. Tactical nuclear weapons are highly destructive and portable, making detection both difficult and essential. Weapons such as dirty bombs can even be hidden among urban environments, making quick detection a matter of national security. Other threats include radioactive isotopes stolen from medical uses, and risks from a malfunctioning nuclear power plant[2].

These threats have driven the development of better technologies for nuclear threat detection. A particular challenge is wide-area detection if radioactive materials make it through choke points such as the borders. For wide area detection, mobile radiation detectors of varying sizes are a promising technology to combat this problem. These mobile radiation detectors can be deployed in a moving vehicle or even carried by pedestrians. In this thesis, we will explore two algorithms which do well depending on sensor size, the first which does well on sensors that can be stored in large vehicles, and the second which does well on sensors which can fit in backpacks.

The task of detecting relatively weak, potentially shielded sources, is particularly difficult when facing variability of patterns of background radiation. This is a common circumstance in cluttered urban environments. It is also in these environments where weapons such as dirty bombs or any improperly stored radioactive material may inflict the most extensive harm. New sensors have made it possible to collect significant amounts of radiation data in real time. Well developed algorithms can aid state of the art sensors in detecting radioactive sources in real time while also maintaining low false alarm rates. A low false alarm rate is of upmost importance; a smoke alarm that beeps every few days even when there is no smoke is useless.

The common approach to detecting potential threat is to subtract the background radiation component from the observed spectral measurements, and judge the probability of the presence of an improperly safeguarded radiation source proportionally to the amount of signal that remains unaccounted by the background model (unsupervised detection). Alternatively, if a spectral template of a particular threat type of interest is known a priori, background suppression can be combined with match filtering of the remaining signal with

the known source template (supervised detection).

In this thesis, we will explore the detection of both known and unknown source types in noisy backgrounds. Specifically, we will explore both algorithms suited for both large sensors and smaller sensors. In the first chapter, we will explore an algorithm known as the G-P Detection Method, which is performs well when larger sensors are used. In the second chapter, we will explore an algorithm named List Mode Regression (LMR) which works well when small sensors are used. Sensors of varying sizes are useful for different situations. Small sensors are weak and noisy, but can be used very close to the source. Large sensors are stronger, but in most cases cannot be put close to the source. For example, if a radioactive isotope exists within a building, then small handheld sensors can be used to locate the whereabouts of the isotope within the building, while the large sensor can only be driven past the building in the back of a van. However, large sensors can cover more ground due to their enhanced detection capabilities, which allows detection at further distances.

The most popular algorithms used with large sensors include Spectral Anomaly Detection with Match Filtering (further labeled as PCA because the anomaly detection part of the algorithms relies on Principal Components Analysis) and Censored Energy Windowing (CEW), a regression-based approach [7][1]. A shortcoming that these solutions share is that they make Gaussian assumptions about underlying data distributions. However, the actual spectral data more closely resembles a Poisson distribution. Hence, we propose two improvements. First, we want to model threat radiation simultaneously with background, and second, we want to model the photon count data as Poisson, but still allow for Gaussian assumptions about the model parameters. We claim that we will be able to capture more subtleties in the inferred background and source spectra that cannot be directly inferred from standard linear analysis. This may be especially crucial in typically low-count observations at higher energies, and when targeted spectra contribute mostly to lower energy bins. We call our model Gaussian Prior on Parameters, Poisson Data Likelihood, or G-P Detection Method (G-P).

With small sensors, total counts is currently the top performing algorithm. We propose a few modifications of CEW which allows us to remove much of the noise that a small sensor injects.

Comparable approaches

In this section we explore existing algorithms used for the task of detecting radiation threats embedded in noisy backgrounds. With larger sensors, two algorithms are currently used. Spectral Anomaly Detection with Match Filtering (PCA) computes a low dimensional principal component projection based on training background formed by multi-energy-bin spectra collected in natural environments without presence of harmful sources [7]. The top components of the PCA model trained on such benign data sample capture the main part of typical variation observed in background-only spectra. Therefore, typical background variation patterns can be subtracted from each future input spectral measurement by dropping these top few PCA components. The resulting residual vector contains spectral discrepancies that cannot be easily explained as usual variation of background. This residual is then correlated with the known source spectrum template vector to produce a source-type-specific score.

Censored Energy Windowing (CEW) method uses the source spectrum template and mean background spectrum vector to determine a subset of energy bins that maximize signal to noise ratio for the specific source type. Then, it uses the energy bins outside the subset as inputs in the regression model trained to predict the sum of photon counts observed in the source-specific energy window under null hypothesis of no presence of a source. More precisely:

$$\hat{B} = (X^T X)^{-1} X^T y$$

Here X is a $N \times |x^C|$ matrix of training samples, where x is the subset of bins that maximize SNR. y is a $N \times 1$ vector of total counts within x. Now given a test spectrum measurement M, let $|M_x|$ be the sum of photon counts with the energy bins that maximize SNR, and M_x^C be the $|x^C| \times 1$ vector of bins outside the maximal SNR bins. The SNR score produced by CEW is $SNR = \dfrac{|M_x| - \hat{B}^T M_x^C}{\hat{B}^T M_x^C}$

The issue with both of these models is that they make Gaussian assumptions about the data, PCA assumes that a PCA projection of a measurement is the best background subtraction, and CEW assumes that a linear combination of bins is the best predictor of background for some subset of bins. Both of these methods also estimate background without considering the shape of the assumed source spectrum being detected.

With smaller sensors, both algorithms fall to a far simpler one, total counts. This algorithm simply sums all photon counts received regardless of energy level, and uses the sum as it's score metric. Total counts performs much better with smaller sensors because the energy readings that a small sensor gives are extremely noisy, so any threat spectra contained within a measurement will also be extremely noisy and virtually indistinguishable from background.

Chapter 1

G-P Detection Method

1.1 Methodology

Given training spectral background measurements X, we model background spectral rates $f = (f_i)_i$ as being Gaussian with mean equal to the mean μ of X and covariance equal to the empirical covariance matrix K of X. We assume a known threatening source spectrum $S = (s_i)_i$ is given, and an unknown source intensity parameter I is Gaussian-distributed with known mean μ_I and covariance σ_I^2. Then, given a spectral measurement y, the log-posterior of background rates f and source intensity I given y and our model is:

$$L(f, I \mid y) \propto -\frac{1}{2}(f-\mu)^T K^{-1}(f-\mu) - \frac{(I-\mu_I)^2}{2\sigma_I^2} - \sum_i (f_i + I s_i) + \sum_i y_i \log(f_i + I s_i)$$

assuming the Gaussian priors on f and I and assuming the Poisson data likelihood for y given f and I [10]. We maximize the log-posterior (over f, I) assuming the alternate hypothesis that a source is present with the given mean and standard deviation on source intensity parameter I, and, separately, also under

the null hypothesis that there is no source present, i.e. $\mu_I = 0$. Furthermore, we enforce that our Poisson parameters remain positive so that they are valid during maximization. Then the difference of these two log-posteriors forms the so-called *log-odds ratio*,which represents the relative odds of the alternative hypothesis (source present) being true versus the null hypothesis (no source present) being true. Strictly speaking, we should integrate the posteriors over all choices of f, I but this proves to be computationally infeasible given the Poisson data likelihood, so we use the maximized posterior probabilities instead for the log-odds ratio calculation. Due to the log-concave nature of the log-posterior, maximization can be performed efficiently using e.g. Newton's method, allowing computations in or under real-time for 1 Hz measurements on a multi-core CPU.

The above procedure applies in a *fully supervised* detection scenario, in which the anticipated type of the radiation sources is known beforehand. We also propose a *weakly supervised* variant of our approach that does not need to rely on the prior knowledge of the specific source type, but instead has access to potentially a very large library of various source templates. This method aims to work when we do not have the exact knowledge of what type of source to expect in the field, but we know that in general it would be similar to one or more templates stored in the library. To accomplish that, we first cluster all known source templates according to their spectral shapes via hierarchical clustering using linear correlation as the distance metric. Now we use the same likelihood as above, except that the background rates f are now the background rates concatenated with the unknown source, and we treat the intensity parameter as fixed to maintain the log concavity of our log likelihood function. We optimize for both the background rates and the source spectrum type at the same time, fixing the intensity parameter I. We use the mean of the source spectra cluster as the source mean and the covariance matrix K is simply a matrix where the upper left quadrant is the empirical covariance matrix for the background and the bottom right quadrant is the diagonal matrix with the variances between the threat templates in the cluster on the diagonal. We compute the log odds ratio in the same way as in the known source case explained above, and we traverse down the branches of the hierarchical clustering tree in pursuit of the greatest ratio.

1.2 Experiments

Our background radiation data was collected using a double 4x16 NaI planar detector driven in the back of a truck in Sacramento, California. Over a period of five days, more than 70,000 one-second observations were collected. An example background measurement is show in Figure 1.1. These observations included counts in 128 energy bins as well as GPS latitude, longitude, orientation, and speed. Noisy GPS readings were removed to yield 50,000 final data points, which were then split into geographically distinct subsets for cross-validation in our machine learning experiments. In addition to the background data, we also have 899 source spectra templates obtained through high fidelity simulations of various configurations of shielded radioactive materials.

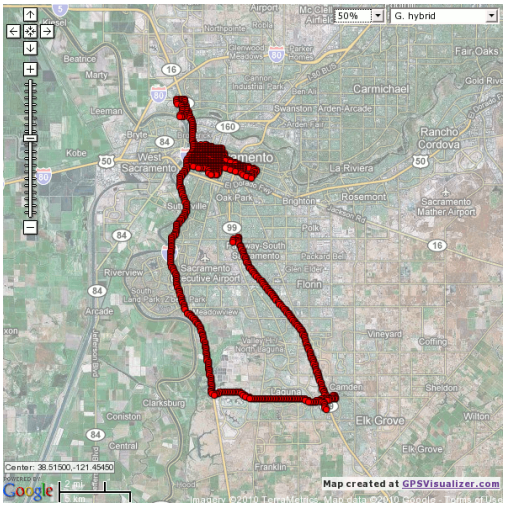


Figure 1.1: Example background measurement

Our synthetic radiation data was generated from background spectra collected in the field (B), as well as source spectra (S). We identify the subset of spectra energy bins, M, that maximize signal-to-noise ratio (SNR) for each source template [6]. We then generate intensity values proportional to $\sqrt{c}$, where $c = \sum_{i \in M} \bar{B}_i$ Intensities I of injections are normally distributed using three different settings of mean = $\{5\sqrt{c}, 10\sqrt{c}, 20\sqrt{c}\}$ and standard deviation = $1/4^{th}$ mean. We do this because it provides a more accurate representation of true SNR than relying on gross photon counts to drive source injections. Each semi-

synthetic true positive data point is therefore obtained as

$$B + \text{Poisson}(I * S)$$

whereas our negative data is simply B collected in the field.

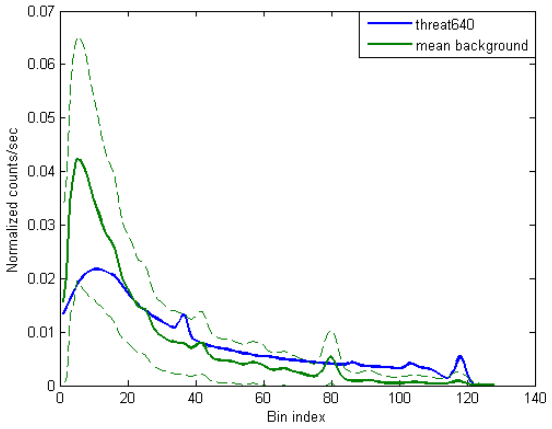
1.2.1 Detectability of sources of particular types

In this section we will present the Receiver Operating Characteristic curves obtained for of our methods compared with PCA and CEW. We choose to present our data on the log scale along the false alert rate axis to highlight the differences in true positive rates at very low false alarm rates, which is crucial for any algorithm that is to be used in real time radiation threat detection. We applied each source template to 20 independently generated sets of positive (source-injected background) and negative (background only) data, in which each data set contains 3,253 true positive data points and 3,253 potential false positive data points, and then calculated log ROC curves with 95% confidence intervals. The first source spectrum template that we used is shown below in Figure 1.3a. Wee see that it is mostly within the 95% confidence interval of the mean background except for high energy bins, which makes this source rather easy to detect. The ROC curve corresponding to this source is also shown in Figure 1.3b (with 95% confidence intervals shown dotted). We see that our method is superior at all false alarm rates against both alternatives, and that on this source, PCA outperforms CEW. Area under ROC curves (AUC) scores shown in Figure 1.2 confirm these observations.

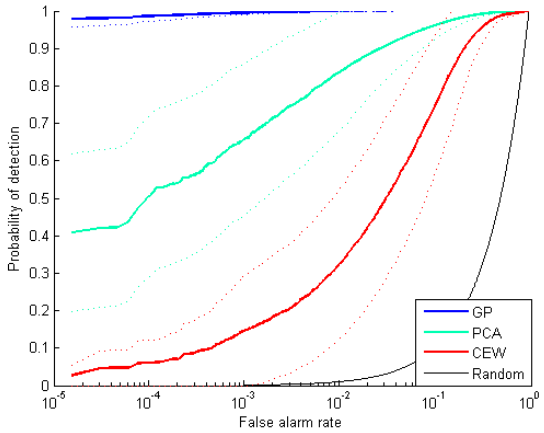
Source	G-P	PCA	CEW
Fig. 1	0.9999	0.9857	0.9225
Fig. 2	0.9999	0.9942	0.9983
Fig. 3	0.8064	0.7836	0.7538

Figure 1.2: AUC scores of each source

Another example of a source where G-P outperforms both PCA and CEW is shown below. However, on this source, CEW does better than PCA. We see that this source spectrum is highly correlated with the background, except at one peak in which the source is outside of the 95% confidence interval of the



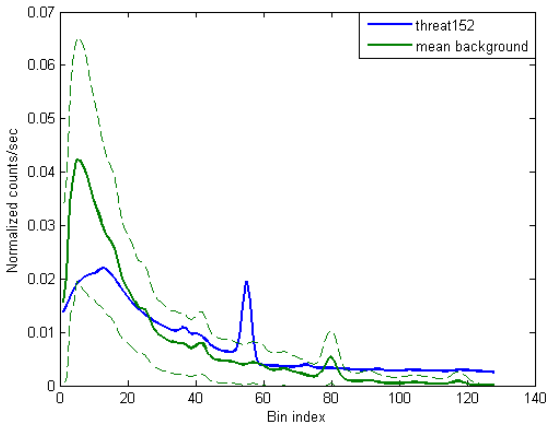
(a) Source spectrum



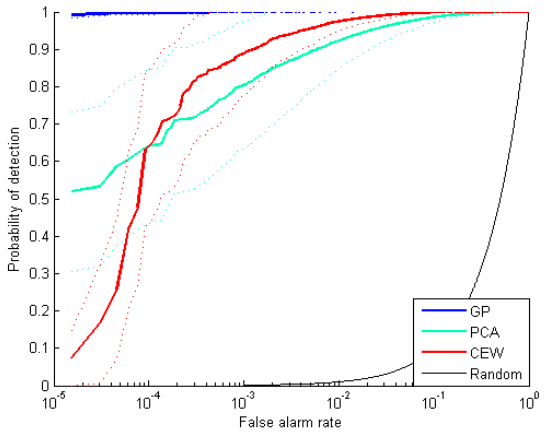
(b) Log ROC curve of source

Figure 1.3: First source spectrum

background. The source spectra and ROC curve is shown in Figure 1.4. Once again, area under curve (AUC) figures which are shown in Figure 1.2 confirm our results.



(a) Source spectrum

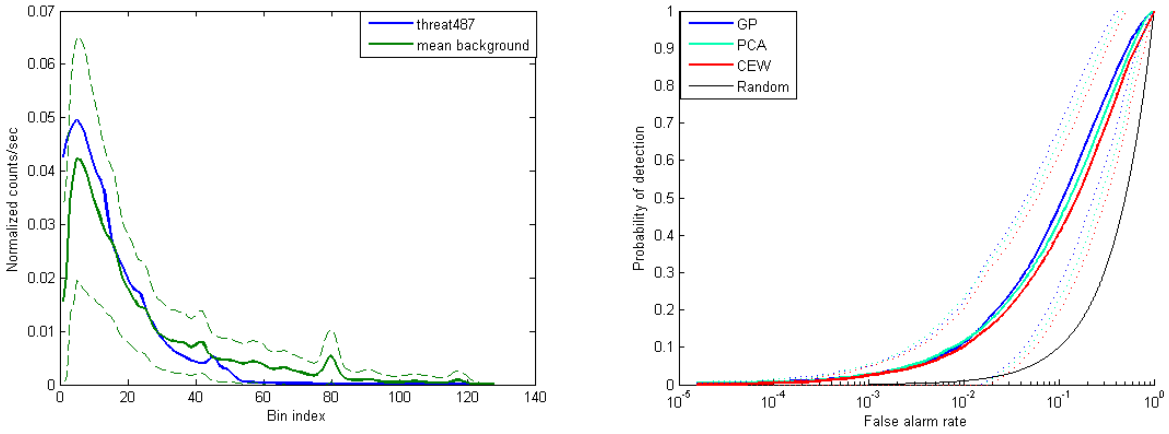


(b) Log ROC curve of source

Figure 1.4: Second source spectrum

Finally, we present a source where neither of the considered methods do particularly well. Our method is still in the lead, although the difference is insignificant. This source template and log ROC plot is shown in Figure 1.5. We see that the source falls well within the 95% confidence interval for all energy bins, and there are no obvious peaks of photon count distribution at any energy interval, which makes this source

extremely hard to detect, even at high intensity. It shows all methods performing poorly at false alarm rates < 0.1. AUC score shown in Figure 1.2 show rather modest separability of negatives and positives in this case.



(a) Source spectrum

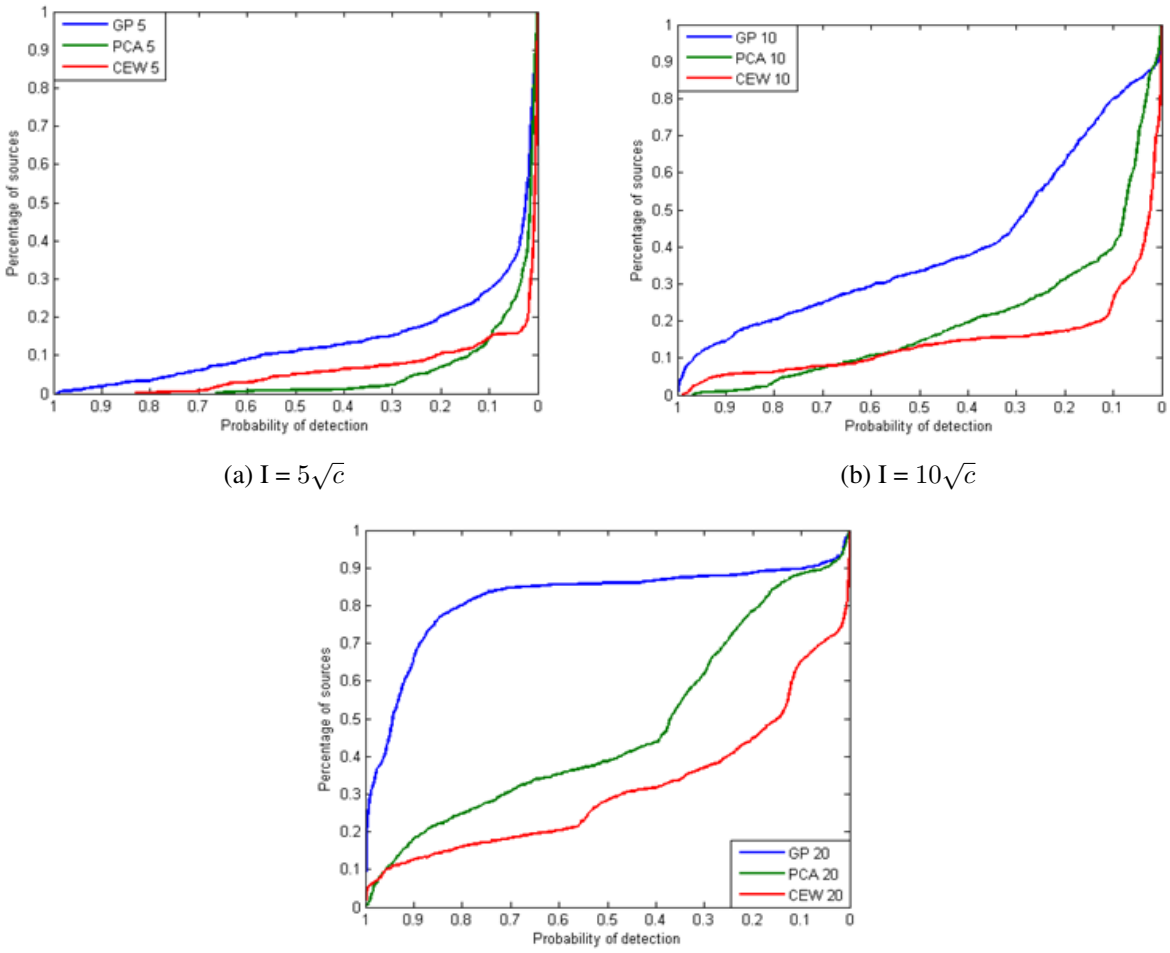
(b) Log ROC curve of source

Figure 1.5: Third source spectrum

1.2.2 Cumulative results over multiple source types

ROC analysis is useful in empirically quantifying detectability of one specific source template at a time. Her we summarize results obtained with all of the available 899 templates. Many of them are very similar, but they all differ slightly, such as in the amount of shielding used. Below are three cumulative distribution functions of probability of detection across all 899 source types for fixed false alert rate of 0.001. Each source type was run with 3,253 independently generated true and potentially false positive data points. Figure 1.6. shows the performance at low, medium, and high intensity levels. We see that G-P outperforms all methods at all intensities, and as intensity increases the margin of victory increases for G-P. Note that G-P at $10\sqrt{c}$ performs almost just as well as PCA at $20\sqrt{c}$. However, these graphs are cumulative distributions, not pairwise comparisons. In the next section we will present results that show pairwise differences between the methods under consideration.

The mean probabilities of detection for each algorithm across all 899 source types are shown in Figure 1.7.



(a) I = $5\sqrt{c}$

(b) I = $10\sqrt{c}$

(c) I = $20\sqrt{c}$

Figure 1.6: Cumulative performance over all source types at I = $5\sqrt{c}$, $10\sqrt{c}$, and $20\sqrt{c}$.

Source intensity	G-P	PCA	CEW
$5\sqrt{c}$	.0243	.0154	.0061
$10\sqrt{c}$	.2779	.0762	.0240
$20\sqrt{c}$	.9428	.3695	.1481

Figure 1.7: Mean probabilities of detection

The proportion of sources types detectable with probability greater than 50% are shown in Figure 1.8

Source intensity	G-P	PCA	CEW
$5\sqrt{c}$	.1112	.0078	.0512
$10\sqrt{c}$	.3348	.1457	.1335
$20\sqrt{c}$	.8587	.3882	.2814

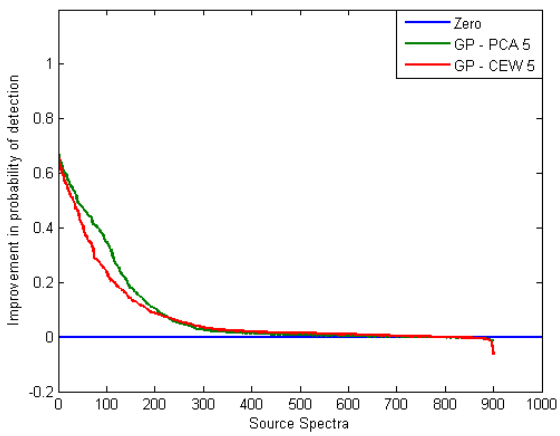
Figure 1.8: Proportion of source with probability of detection $> 50\%$

In both instances, our method outperforms both competitors by a significant margin. Running a paired two sample t-test between G-P vs. PCA and G-P vs. CEW reveals that all p-values are less than 10^{-50}, indicating a very high significant difference.

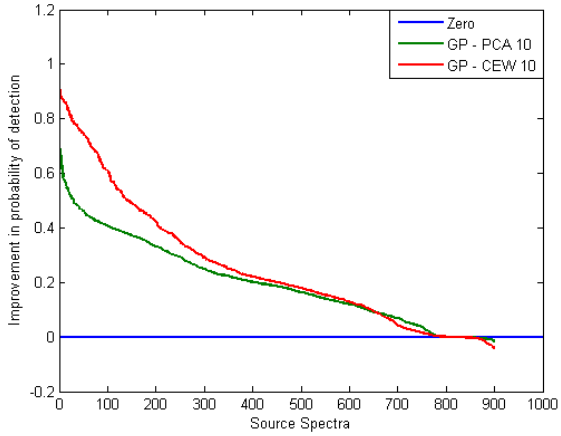
1.2.3 Pairwise analysis

Figure 1.9 shows the pairwise differences of probability of detection at fixed false alarm rate of 0.001 between the three methods at low, medium, and high intensities. As the plots show, G-P wins on a majority of sources at all intensity values, and the proportion of sources on which it is victorious increases as the intensity increases. The exact proportions of victories for G-P are shown in Figure 1.10.

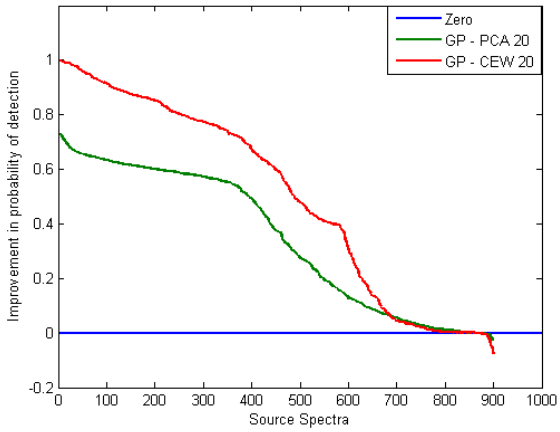
We see that even at high intensity, PCA beats our G-P method on 50 out of 899 sources. We claim that these sources are very hard to detect for both methods, and the reason why PCA does better than G-P is pure chance. To validate this claim, we independently repeated data generation 20 times independently, for each of those 50 source templates. We then compared probabilities of detection at false alarm rate of 0.001. Figure 6 depicts the proportion of times the PCA method outperforms our G-P for the 50 sources in question, with 95%-ile confidence intervals. G-P systematically beats PCA on 43 out of the 50 sources with 95% confidence. Only one source out of 50 had PCA beating G-P more than 50% of the time, but that result



(a) I = $5\sqrt{c}$



(b) I = $10\sqrt{c}$



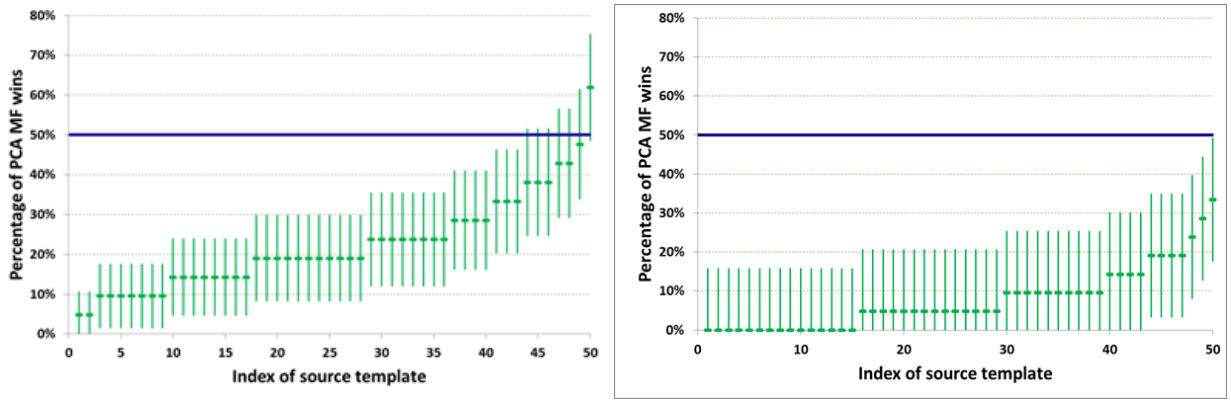
(c) I = $20\sqrt{c}$

Figure 1.9: Pairwise difference of G-P vs. PCA and G-P vs. CEW in terms of probability of detection at fixed false positive rate 0.001

Source intensity	G-P vs. PCA	G-P vs. CEW
$5\sqrt{c}$	.8109	.8676
$10\sqrt{c}$	.8943	.8921
$20\sqrt{c}$	.9444	.9611

Figure 1.10: Proportion of GP victories

was not significant at 95% confidence. We conclude from this experiment that PCA does not systematically beat G-P on any sources, but instead by chance for hard to detect sources in which neither algorithm does particularly well. Also, to further validate our claim, we conducted the same comparison for the 50 sources where G-P won by the lowest margin. The results shown in Figure 1.11 indicate significant victory rates for all those 100 sources at 95% confidence.

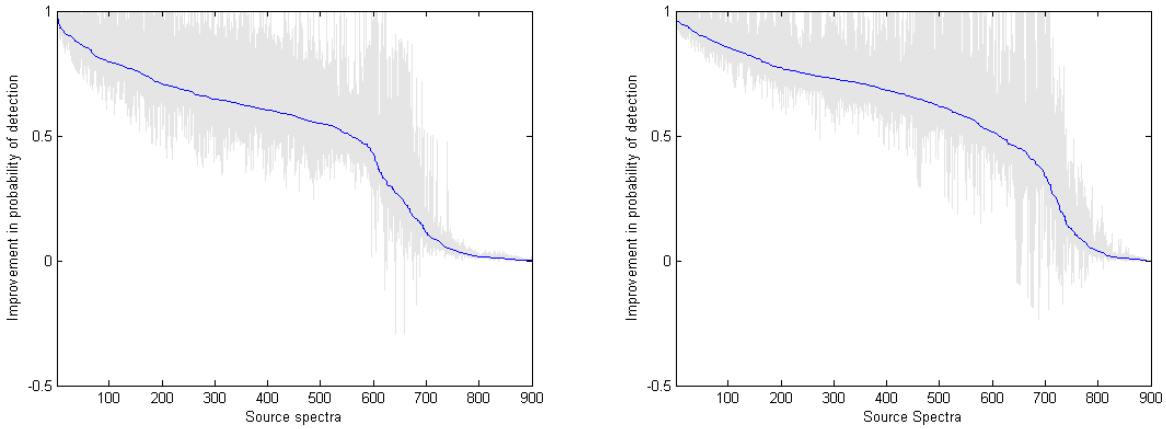


(a) Reruns of 50 sources which PCA outperformed G-P (b) Reruns of 50 sources which G-P outperformed PCA by the lowest margin

Figure 1.11: 20 independent runs of sources which PCA performed closest to GP on

1.2.4 Re-run of all sources

In order to establish that the previously established results were systematic, we performed pairwise bootstrap analysis for all of the source templates in our library. Plots in Figure 1.12 show the mean improvement in probability of detection at false positive rate of 0.001 for all 899 source templates sorted decreasingly, respectively for G-P vs. PCA and G-P vs. CEW. The graphs include 95%-ile confidence bands obtained from 20 independently generated sets of data. These results validate previously observed dominating performance of the method proposed in this paper over the more familiar alternatives.



(a) Pairwise difference of GP - PCA

(b) Pairwise difference of GP - CEW

Figure 1.12: Pairwise difference of probabilities of detection at false alarm rate of 0.001 over the entire threat library

1.2.5 Detection of sources of unknown types

Here, we present results from running our weakly supervised G-P model to detect sources of radiation of unspecified types. In this case, our model simultaneously optimizes for both the background rates and the source spectrum template. We chose 50 sources randomly from our template library and generated synthetic radiation data in the same way as before. We then attempt to simultaneously detect threat and infer its most likely spectrum, by testing hypotheses stored in the dendrogram of a hierarchical clustering model estimated from the source template library. Each node of the dendrogram corresponds to a cluster of source templates that can be tested for the goodness of fit with the observed spectral measurements. Our algorithm iteratively traverses the hierarchical clustering tree in order to find the node that yields the highest log likelihood ratio. This node corresponds to the cluster of templates that to our algorithm appears to be the most likely to include the sought after unknown source type.

In order to validate relevance of thusly identified likely matching source types, we calculate the correlations of each of the source templates in the final cluster with the original observation after background photon counts have been removed from it, and select the most correlated template. Apparently, many source templates are highly correlated: the average correlation between all their pairs is 0.92. In this case, even very slightly higher than average correlations may indicate significant matches. We therefore calculated the correlations of all the templates with the background-suppressed measured spectrum and then found the

rank of the template found as the most likely match in the winning cluster. It was 0.98 while the top ranking score across the complete library was 0.99. The median rank across the conducted batch of experiments was 62 out of 899.

In an alternate scenario, the source that we are supposed to detect may not be represented within our library at all. To test the resiliency of our G-P method to this kind of adverse scenario we randomly removed a subtree from the hierarchical clustering model and all library templates stored in it. Then, we repeated the same experiment as before, running each source we removed through the G-P method. The median rank of the most correlated source in the cluster we found was 85 out of the 888 remaining source templates, and the observed correlation was 0.97.

In Figure 1.13, we show the ROC curves for the first source type used as an illustration in this paper including the result of the hierarchical clustering based G-P method, after the cluster of size 8 containing this very template has been removed from the tree. We see that the hierarchical G-P method does not do as well as G-P or PCA, but all alternatives have prior knowledge of the exact source template being sought. We see that even without the knowledge of the source spectrum, hierarchical G-P degrades gracefully and it still can outperform CEW (which knows the source) at a range of settings.

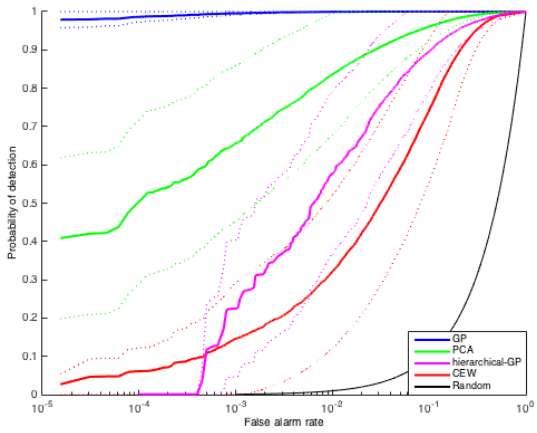


Figure 1.13: ROC for the source from Fig.1 including the hierarchical G-P model.

1.3 Conclusion

We have shown that our G-P method makes a highly competitive tool for complex tasks such as detection of faint radioactive sources embedded in noisy backgrounds. It typically and most often very significantly outperforms current commonly used algorithms including Spectral Anomaly Detection with Match Filtering and Censored Energy Windowing. In cases when the exact spectral templates of the sources being sought are not known a priori, G-P shows promise to degrade gracefully due to using weak supervision via hierarchical clustering of a large library of source templates. By modeling threat radiation simultaneously with variable background, and modeling the radiation data as Poisson with Gaussian assumptions on model parameters, we are better able to capture subtleties in the inferred background, allowing more reliable detection of potentially harmful sources of radiation.

Future work includes testing our G-P method using data obtained with smaller spectrometers, scenarios involving large speed of relative motion of a sensor and the source, and at large standoffs between the source and the sensor. All these cases involve low expected photon counts either from just the source, or both from background and sources, where using the correct Poisson assumptions on data should yield the most substantial benefits [14]. We hope to identify boundaries of the beneficial use of the G-P method in order to substantially improve the effectiveness and range of usability of intelligent radiation threat detection systems.

Chapter 2

List Mode Regression

2.1 Background

Suppose there exists some dirty bomb which is hidden in an urban setting. A proposed defense against this sort of attack is to equip personnel with backpack sized sensors. These agents can then walk around in a grid like pattern, while their sensors transmit measurement information to a centralized server. This will allow fast detection of a potentially harmful radioactive source within an urban environment.

The use of sensors of this size presents new challenges. The total number of counts absorbed is around 1/25th of what the 4x16 NaI planar detectors we used in the previous section absorbed. Low total photon counts gives us less information to detect if a measurement contains source spectra. The Compton Effect is also much stronger in a small sensor which injects more noise in the measurement of a single photon's energy[3]. The Compton Effect causes higher energy photons to be measured as lower energy photons by a sensor. This further exacerbates the challenges of dealing with low total counts.

Due to the strong Compton Effect, the current state of the art algorithm used to detect source with small sensors is total counts. This algorithm simply sums all photons regardless of energy and uses the sum as a score metric, i.e $score = len(l)$, where l is a list of photons for each measurement. Previous methods that we discussed such as G-P, PCA, and CEW all perform poorly on these sensors due to noise caused by the Compton Effect, with very low probabilities of detection at acceptable false alarm rates. G-P does poorly due to the fact that the covariance matrix contains little information due to the nature of low counts. This is

show in Figure 2.1.

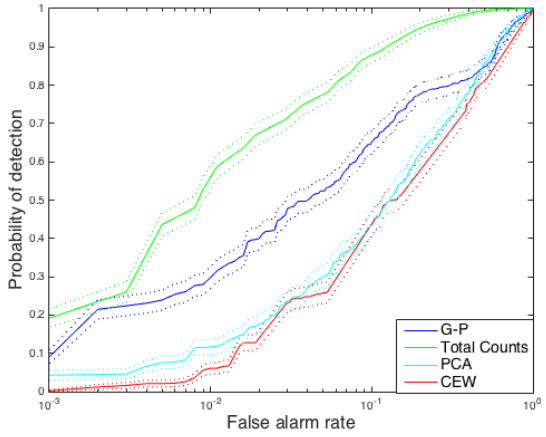


Figure 2.1: Log ROC of G-P, PCA, CEW on a small sensor

2.2 Methodology

Let l be a list of photons for each measurement, where each entry in the list corresponds to the energy of a photon received by a gamma-ray spectrometer. Then, given a sensor response matrix, which gives a probability density function of true photon energies given the measured photon energy, each received photon is mapped to a possible true energy. To accomplish this, a Monte Carlo approach is used. Each measured photon is replicated into 1000 identical copies, and then, for each such replica, its possible true energy is randomly drawn according to the probability density function given by the sensor response model. This is done for each measured photon in l to obtain l'. The source spectrum template and mean background spectrum are used to determine a subset of energy ranges that maximize signal to noise ratio for the specific source type. l' is separated into photons that are in the maximal SNR energy ranges and photons that are outside of these ranges. The predicted variable is simply the number of photons inside the maximal SNR energy ranges, y. The photons outside of this range are now a list of photons which follow some distribution P. These distributions are then compared by taking inner products between their mean embeddings in a reproducing kernel Hilbert space, as in [3]. For computational efficiency, the embedding space used is that of random Fourier features [4] approximating the Gaussian RBF kernel, as considered e.g. in [2]. Thus, letting $P(\omega)$ be the Fourier transform on the Gaussian RBF kernel and $z : \mathcal{X} \to \mathbb{R}^D$, distributions are

compared with the Mean Map Kernel as:

$$\mathrm{MMK}_z(X,Y) = \left(\frac{1}{n}\sum_{i=1}^{n} z(X_i)\right)^T \left(\frac{1}{m}\sum_{j=1}^{m} z(Y_j)\right)$$

$$z(x) = \sqrt{\frac{2}{D}} \begin{bmatrix} \cos(\omega_1^T x + b_1) \\ \vdots \\ \cos(\omega_D^T x + b_D) \end{bmatrix}, \omega_i \sim P(\omega), b_i \sim \mathrm{Unif}(0, 2\pi).$$

For more information about the mean map kernel, we direct interested readers to [13]. We use Kernel Ridge Regression with the above mean map kernel to train our model, where we want to minimize

$$J(w) = (y - Xw)^T(y - Xw) + \lambda\|w\|^2$$

The optimal solution is thus

$$w = (X^T X + \lambda I)^{-1} X^T y$$

By using the identity

$$(P^{-1} + B^T R^{-1} B)^{-1} B^T R^{-1} = PB^T(BPB^T + R)^{-1}$$

we can rewrite this as

$$w = X^T(XX^T + \lambda I)^{-1} y$$

We can then define the following dual variables

$$\alpha = (K + \lambda I)^{-1} y$$

and then rewrite the primal variables as

$$w = X^T \alpha = \sum_{i=1}^{N} \alpha_i x_i$$

Now at test time[11]

$$\hat{f}(x) = w^T x = \sum_{i=1}^{N} \alpha_i k(x, x_i)$$

Once we are done with our kernel ridge regression, we can calculate a score from y and $\hat{f}(x)$ for each measurement which will be used to classify if a measurement contains source or not. We use empirical SNR of the observed vs. expected photon counts in window y as our source detection score metric, where

$$SNR = \frac{y - \hat{f}(x)}{\sqrt{\hat{f}(x)}}$$

2.3 Experiments

List mode data used for these experiments were collect by the RadMAP project at Berkeley. The RadMAP data [8] set represents the observations of the MISTI mobile detection vehicle [4] traveling the roads of West Berkeley, Downtown Berkeley, South Side, and Solano California for four hours on May 1, 2012 from 9:25am to 1:45pm. Radiological data was stored in list mode, recording time, energy, and photon count. Total background detected by the NaI array with coded aperture in place averaged approximately 10,000 counts per second. Positional data was measured at a frequency of 5Hz recording time, latitude, longitude, altitude, pitch, and yaw. Average vehicle velocity was 5.4m/s. Figure 2.2 shows the path of the detection vehicle represented in this data set. Observations are represented as black points. The shaded gray region represents the space of candidate source locations. The 100 NaI detectors on the detection vehicle were arranged in a 10×10 grid. The coded aperture was placed 0.4m in front of the detector array [4]. Each detector element measured 0.1m in both height and width.

Source spectra was obtained from a similar dataset that used the same sensors. A small sensor response model was provided to us by Lawrence Livermore, which models the Compton effect. Our setup is as follows. Lists of background counts are generated from the RadMAP data and random poisson draws are used to determine how many counts are contained per measurement. Then source is injected from the East/West dataset for true positive samples. Source counts are generated by rejection sampling of the threat spectra. Rejection sampling works as follows. To obtain a sample from distribution X with density $f(x)$ using samples from distribution Y with density $g(x)$ where $f(x) \leq Mg(x)$ for all x, we first obtain a

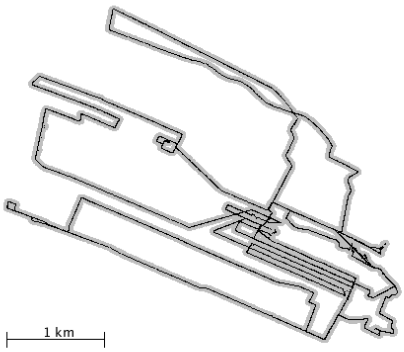
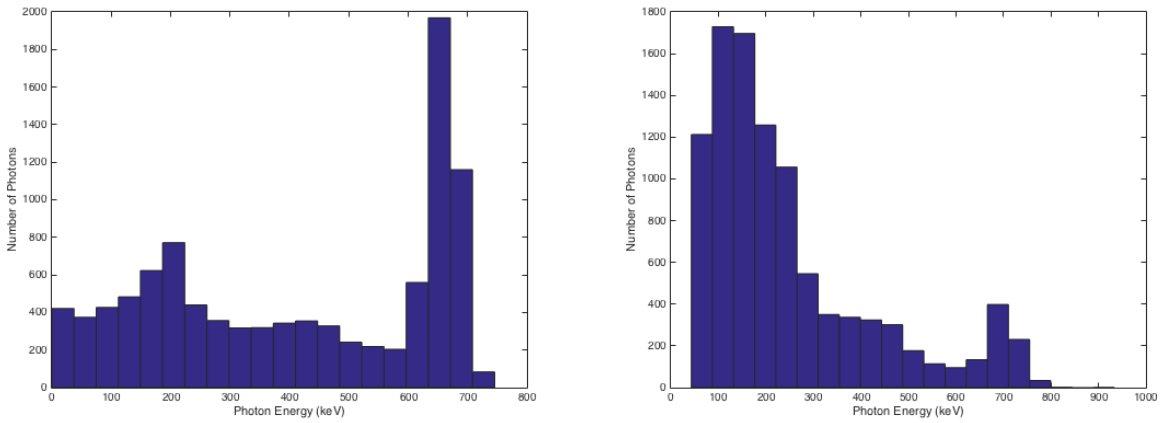


Figure 2.2: Path of RadMAP detection vehicle in empirical data.

sample y from distribution Y and a sample u from $Uniform(0, 1)$. We accept this sample if $u < \frac{f(x)}{Mg(x)}$, otherwise we reject[16]. We then generate a random draw for each photon in the list according to the small sensor response matrix to obtain a 'measured photon'. Lists of 'measured photons' are the base dataset that we use for the following experiments. Small portable sensors average around 60 counts per second. Thus, background samples averaged 60 photons each. To generate synthetic observations, we used the following procedure. For each 20 meter area in the RadMAP dataset, we found the Maximum Likelihood Estimator of the total counts of all the observations in the area, c. We then draw from $Poisson(\alpha\bar{c})$, where $\alpha = \frac{60}{c}$ and $\bar{c}$ is simply the mean of all c, and use the value of that draw as the number of photons in the background sample. This is done to preserve background variation that exists within the original measurements, while generating reasonable small sensor background observations. Source signatures were injected into thusly prepared small-sensor background data with a mean of 15 source-originating photons per measurement to produce an overall expected $SNR = \frac{S}{\sqrt{B}} = \frac{15}{\sqrt{60}} \sim 2$ per measurement.

2.3.1 Detection of known sources from a small sensor

There are three main advantages List Mode Regression has over traditional CEW and Total Counts. The first advantage is that we are able to account for the sensor response model. Each threat spectrum has a peak of photons at a given range of energies. A small sensor response model causes the peak to almost disappear. Also recall from Figure 1.3a that a large peak in low energy photons coincides with background counts.



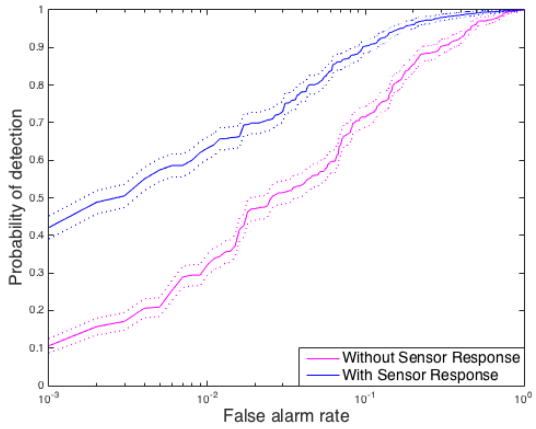
(a) Original threat spectrum

(b) Threat spectrum after sensor response model

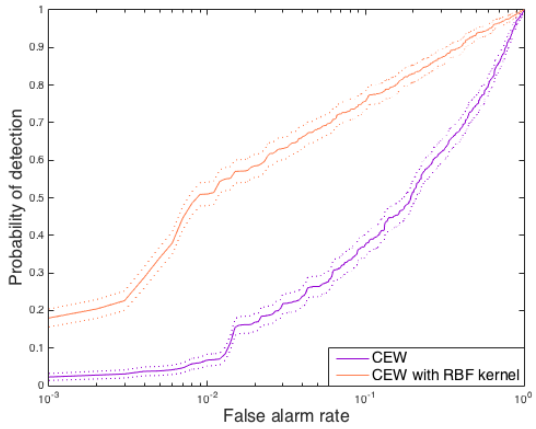
Figure 2.3: Histogram of Cs137 threat spectra

Thus, we want to reconstruct this peak by taking into account the sensor response model. The Compton Effect changes the distribution of photon energies that are measured, and the sensor response model is used to attempt to undo this transformation. If we simply run our threat spectrum through the sensor response model, our maximal SNR ranges include almost all counts, rendering regression useless. Thus, our Monte Carlo approach maps these 'measured photons' from the sensor into 'true photons' that we can use to better regress on. We show this improvement in detection due to the response model in Figure 2.4a. The blue line shows LMR with the sensor response model, and the green line shows LMR without the sensor response model on the same dataset after the sensor response matrix transformation. Furthermore, we show that only our method benefits from the sensor response model. CEW is unable to effectively use the sensor response model because the energy bin ranges are not fine enough, so transformed source and threat will still end up in the same energy bin. Figure 2.4c shows CEW with and without the sensor response model. All of the following log ROC plots in this section used threat spectra Cs137. Cs137 is an industrial threat spectra.

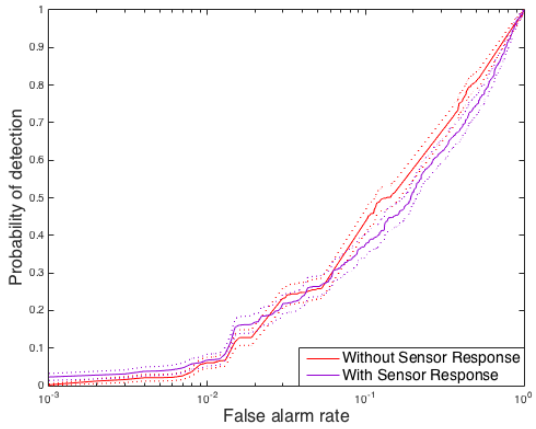
The second advantage LMR has is the ability to use a powerful kernel to leverage non-linearities in the data. However, without the response model LMR is unable to perform better than total counts because of the strong Compton Effect. We show this by comparing traditional CEW with a modified CEW where we use kernel ridge regression instead of normal ridge regression in Figure 2.4b. Both these methods make used of binned data, so we simply rebin our list mode data. Also, both these methods used the sensor response model, so we show that even with a kernel and a sensor response model, CEW falls short.



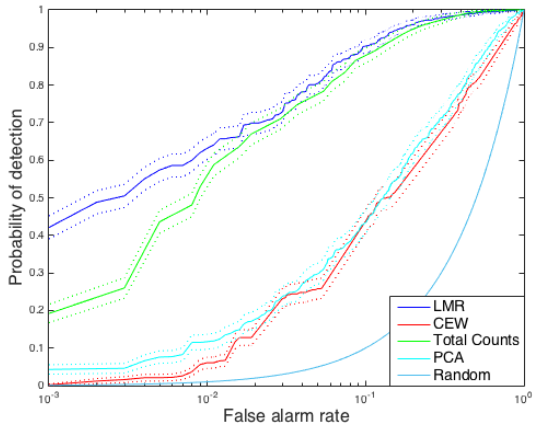
(a) Log ROC of LMR with and without the sensor response model



(b) Log ROC of CEW with and without a kernel



(c) Log ROC of CEW with and without sensor response model



(d) Log Roc of all methods

Figure 2.4: Log ROC's of various methods for spectra Cs137

The third advantage comes from using distribution-based regression, which can be learned directly from their non-parametrically represented energy distribution, rather than learning a standard multivariate regression using binned vector data representation, as in the standard CEW model. The boost in performance here can be seen from the difference between the blue lines in Figure 2.4a and Figure 2.4b. The blue line in Figure 2.4b uses standard multivariate regression, while the blue line in Figure 2.4a uses distribution-based regression.

Finally, we show our full method vs traditional CEW, PCA, and total counts in Figure 2.4d. LMR outperforms total counts by almost 100% at the lowest false alarm rate, which is where we are the most

concerned. Furthermore, LMR shows no dropoff vs total counts even at higher false alarm rates.

2.3.2 Cumulative performance over many source types

To validate our claim that LMR performs better than its competitors, we ran the same experiment over 40 distinct source types. These 40 source types have many uses, some are medical, others are industrial, and some are even fissile. Variances in the performance of total counts is not due to the source type, but rather of the poisson assumption we made while generating background measurements. Figure 2.6 shows the pairwise improvement in probability of detection at a fixed false alarm rate of 0.001. LMR does better on some sources vs Total Counts and worse on others. The mean improvement in probability of detection at false alarm rate of 0.001 is 0.05. Easier to detect sources generally have sharper peaks in their source spectra. A way to measure this is to compute the proportion of the threat spectra that is contained within the maximum SNR ranges used in LMR. Figure 2.5 shows the relationship between the proportion of threat that is contained within the maximum SNR ranges vs the probability of detection using LMR at false alarm rate of 0.001 for all 40 sources. The r^2 value of this relationship is 0.44. We use the probability of detection using LMR as our response variable rather than the improvement in probability per source vs Total Counts since the performance of Total Counts does not depend on which source spectra is used.

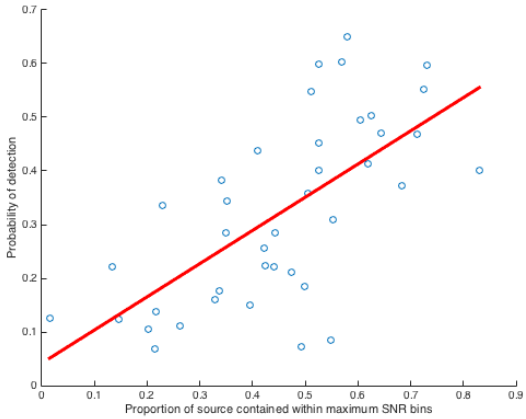
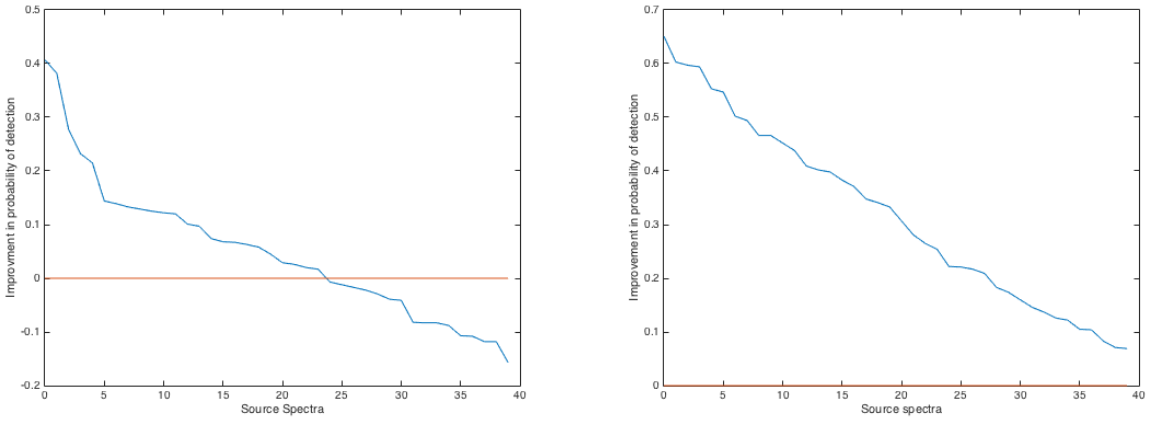


Figure 2.5: Probability of detection at false alarm rate of 0.001 vs proportion of threat that is contained within the maximum SNR ranges



(a) LMR - Total Counts

(b) LMR - CEW

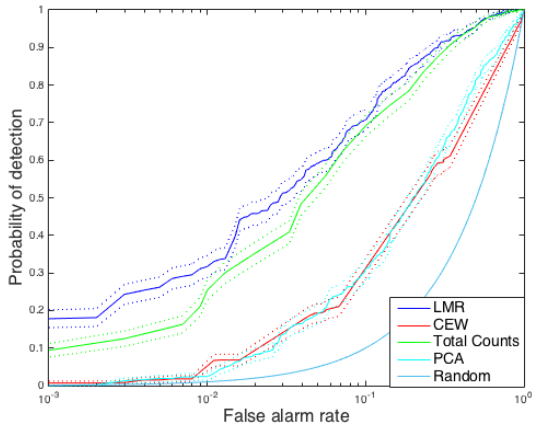
Figure 2.6: Pairwise improvement in probability of detection of LMR vs competing methods over 40 source spectra

2.3.3 Variation in background and SNR

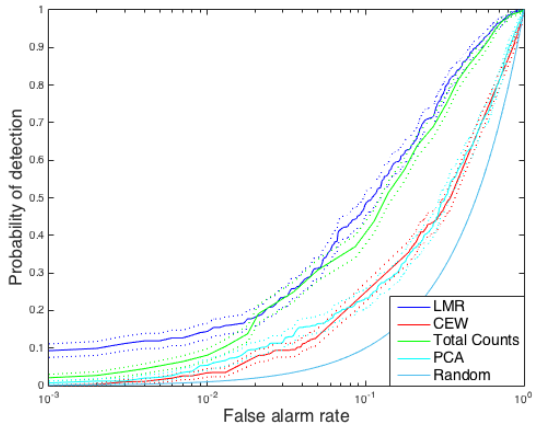
We also wanted to test the performance of LMR at various background intensity settings and SNRs. This is equivalent to varying the size of the sensor we are using, and how strong the source we are detecting is. Our motivation for this experiment is to show LMR's robustness at low counts, and that LMR's improvement in performance is not specific to our experimental setup. Figure 2.7a shows that LMR continues to beat total counts at a lower background count of $Poisson(30)$ while holding SNR constant. Furthermore, Figure 2.7b shows that at the lower background rate with $SNR = 1.3$, LMR still outperforms total counts. Using slightly larger sensors with a background count of $Poisson(120)$, we also see an improvement in detection capabilities.

2.4 Conclusion

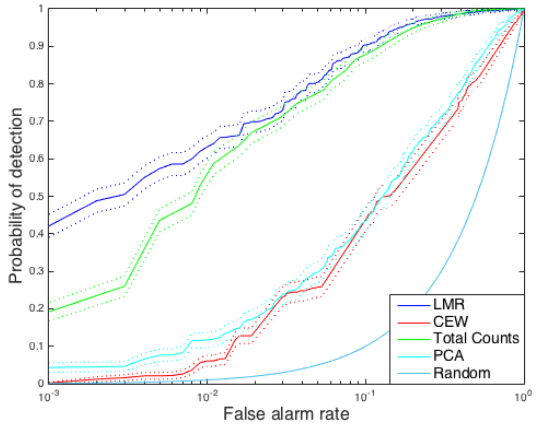
From these experiments we observe that LMR approach offers noticeable advantages at detecting radiological threats in low-particle-count scenarios. Specifically, the ability to learn from distributions of photon energies without having to bind them into energy intervals, and the use of the sensor response model, allow extending the range of usability of Censored Energy Window methodology to low-count applications. It allows for accurate and source-type-specific detection even if total background photon count rates are in the



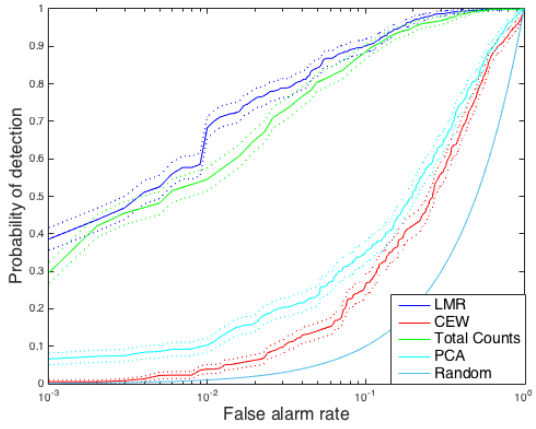
(a) Background $\sim Poisson(30)$ with $SNR = 2$



(b) Background $\sim Poisson(30)$ with $SNR = 1.3$



(c) Background $\sim Poisson(60)$ with $SNR = 2$



(d) Background $\sim Poisson(120)$ with $SNR = 2$

Figure 2.7: Log ROC curves of Cs137 at varying background rates and SNRs

range of tens per second, and if the source signatures do not exceed signal-to-noise ratios of 2. We see an improvement in detection capabilities for the majority of considered source spectra.

Future work includes merging this algorithm into the Bayesian Aggregation framework described by Tandon[15]. A hypothetical use case is as follows: workers with portable sensors transmit data to a centralized server. The centralized server is then able to aggregate these observations with the Bayesian Aggregation framework to better detect source spectra. We would also like to repeat the bootstrap analysis done with the G-P method here. To further prove robustness of LMR, we would run LMR with an inaccurate sensor response matrix and evaluate performance. Further optimization of LMR's source code will allow us to evaluate it's performance on the 4x16 NaI sensors used in Chapter 1. Currently LMR is too computationally

expensive to use for these larger sensors.

Bibliography

[1] Christopher Bishop. *Pattern Recognition and Machine Learning*. Springer Publishing Company, Incorporated, 2006.

[2] *Fukushima accident*. URL: http://www.world-nuclear.org/information-library/safety-and-security/safety-of-plants/fukushima-accident.aspx.

[3] Glenn Knoll. *Radiation Detection and Measurement*. Wiley, 2010.

[4] Lee J Mitchell et al. "Mobile imaging and spectroscopic threat identification (MISTI): system overview". In: *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*. IEEE. 2009, pp. 110–118.

[5] Krikamol Muandet et al. "Learning from Distributions via Support Measure Machines". In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 10–18. URL: http://papers.nips.cc/paper/4825-learning-from-distributions-via-support-measure-machines.pdf.

[6] Karl Nelson and Simon Labov. *Aggregation of Mobile Data*. Tech. rep. Lawrence Livermore National Lab, 2012.

[7] Karl Nelson and Simon Labov. *Detection and alarming with sords unimaged data: Background data analysis*. Tech. rep. Lawrence Livermore National Lab, 2009.

[8] Brian Quiter, Lavanya Ramakrishnan, Mark Bandstra, et al. *GRDC: a Collaborative Framework for Radiological Background and Contextual Data Analysis*. Tech. rep. Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA (United States), 2015.

[9] Ali Rahimi and Benjamin Recht. “Random Features for Large-Scale Kernel Machines”. In: *Advances in Neural Information Processing Systems 20*. Ed. by J. C. Platt et al. Curran Associates, Inc., 2008, pp. 1177–1184. URL: http://papers.nips.cc/paper/3182-random-features-for-large-scale-kernel-machines.pdf.

[10] Carl Edward Rasmussen and Christopher Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. Tech. rep. Cambridge, Massachusetts, 2006.

[11] Craig Saunders, Alexander Gammerman, and Volodya Vovk. “Ridge Regression Learning Algorithm in Dual Variables”. In: *Proceedings of the Fifteenth International Conference on Machine Learning*. ICML ’98. 1998, pp. 515–521. ISBN: 1-55860-556-8. URL: http://dl.acm.org/citation.cfm?id=645527.657464.

[12] Alexander Smith. *What to Know About Nuclear Weapons in 2015*. URL: https://goo.gl/g2Zt8j/.

[13] Dougal J. Sutherland and Jeff G. Schneider. “On the Error of Random Fourier Features”. In: *CoRR* (2015). URL: http://arxiv.org/abs/1506.02785.

[14] Prateek Tandon. “Bayesian Aggregation of Evidence for Detection and Characterization of Patterns in Multiple Noisy Observations”. PhD thesis. Carnegie Mellon, 2015.

[15] Prateek Tandon et al. “Detection of radioactive sources in urban scenes using Bayesian Aggregation of data from mobile spectrometers”. In: *IS* (2016). URL: http://www.sciencedirect.com/science/article/pii/S0306437915001866.

[16] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Publishing Company, Incorporated, 2010, p. 421.