Non-negative Matrix Factorization of Gamma-ray Spectra for Background Modeling, Detection, and Source Identification

K. J. Bilton, T. H. Joshi, M. S. Bandstra, J. C. Curtis, B. J. Quiter, R. J. Cooper, and K. Vetter

Abstract—**Radiological source search is a challenge involving the detection, identification, and localization of weak sources within background environments that vary both spatially and temporally. In this work, a method for simultaneously detecting and identifying gamma-ray sources using background models formed from spectral data is described. Non-negative matrix factorization (NMF) is used to generate low-dimensional representations of gamma-ray spectra, allowing for a compact means of capturing variation in gamma-ray backgrounds. Background models formed using NMF are used to perform anomaly detection, and additionally, models are augmented with spectral templates of gamma-ray sources to perform simultaneous detection and identification using a likelihood ratio test. The NMF-based detection and identification algorithm is benchmarked against a standard Region of Interest algorithm and shows significant performance gains. Additionally, NMF-based anomaly detection shows improvements over methods based on gross counts or Principal Component Analysis. Algorithm performance is evaluated using unshielded sources with activities between 5 μCi and 400 μCi at a standoff distance of 20 m using source injection on background data collected using a 1 m^2 NaI array on the Radiological Multi-sensor Analysis Platform.**

Index Terms—**Gamma-ray detection, gamma-ray spectral analysis, radiation source search, anomaly detection, non-negative matrix factorization**

I. INTRODUCTION

DETECTING and identifying radiological sources within environments with dynamic radiation backgrounds is essential for applications in homeland security, emergency response, environmental monitoring, and contamination remediation. In particular, radiation source searches involve detecting and identifying potentially weak sources in complex background radiation environments by means of mobile

radiation detection systems. In most cases, rates and energy distributions of background gamma radiation are not known *a priori*, making it challenging to know what should be considered an anomalous measurement, particularly for weak sources.

Alongside difficulties associated with the radiological background, there are operational constraints which limit the effectiveness of anomaly detection algorithms. In practice, false alarms may elicit a response from system operators, and as a result, the acceptable false alarm rate (FAR) (e.g., 1 alarm in 8 hours) places a lower limit on decision thresholds, affecting the detection sensitivity. Furthermore, medium-resolution scintillation detectors (e.g., NaI) are often used, primarily due to lower cost per unit volume, and in turn, the most prominent spectral features from a radiological anomaly typically span a relatively broad range of energies.

A variety of methods are used to detect and identify radiological sources, some of which mitigate the previously described challenges by accounting for the variability seen in the radiological background. Among the simplest approaches are a class of algorithms that only consider the measured gross count rate, neglecting spectral information and only alerting operators when the total counts exceeds a threshold. As a result of neglecting spectroscopic information, such algorithms lack the ability to identify the gamma-ray source causing the alarm. Additionally, background count rates vary drastically, and methods based on the gross count rate require high thresholds to maintain a reasonable FAR. More sophisticated approaches incorporate spectral information, and include methods that estimate the number of counts in a region based on nearby windows or consider ratios of counts within different spectral regions [1]. Even more advanced algorithms use the shape of the entire spectrum to identify deviations [2], [3], [4], or may aggregate test statistics from sequential measurements to exploit time dependence. The reader is referred to [5], [6], [7] for more details and comparisons between various methods.

The methods introduced in this paper make use of the entire gamma-ray spectrum, and we begin by discussing the advantages and disadvantages of different full-spectrum decomposition methods used for anomaly detection and identification. One such method is noise-adjusted singular value decomposition (NASVD), a spectral decomposition technique developed for airborne gamma-ray measurements to reduce noise in spectral bins due to Poisson sampling [8], [9]. NASVD removes linear combinations of spectral features associated with small singular values, and then transforms data

back into the original dimensionality, yielding filtered spectra. The method implicitly assumes Gaussian statistics, and the resulting spectral components are orthogonal under the inner product. Although the first several components may resemble spectra from known background contributions such as ^{40}K and uranium and thorium series radioisotopes, the orthogonality of the components rules out true physical solutions, which must be non-negative. Nevertheless, NASVD has been successfully used for detecting spectral anomalies and extracting weak signals from airborne measurements [10] and for developing robust real-time background estimators [11].

A nearly identical technique to NASVD is Principal Component Analysis (PCA), which has been used on portal monitor data [3], among other types of detection systems. It is able to detect and identify sources in transit past portal monitors, but like NASVD, its background decomposition cannot have a consistent physical interpretation.

More recently, Poisson Principal Component Analysis (PPCA) has been used to accurately model the underlying Poisson statistics of photon detection [12]. PPCA is a specific case of Exponential family Principal Component Analysis [13], and it has shown an improvement over standard PCA in source detection at low signal-to-noise, presumably due to more accurately modeling the statistics at low counts [14]. Although PPCA assumes the correct statistical model and generates non-negative spectral components, the components still lack an intuitive, physical interpretation because they are defined in logarithmic space and are therefore multiplicative, not additive. While PCA, NASVD, and PPCA often result in models containing features that one may recognize as resulting from physical processes (e.g., a component containing a 1460 keV peak from ^{40}K), these methods do not fully capture the additive, non-negative nature of gamma-ray spectra.

Non-negative matrix factorization (NMF) is a dimensionality reduction technique that does not suffer from the drawbacks associated with PCA and PPCA [15], [16]. NMF has been applied to many problems, including facial recognition [16], [17], textual analysis [16], and particulate pollution analysis [18]. Because it models non-negative linear mixtures, NMF has been used successfully in many spectroscopic applications, including magnetic resonance chemical shift imaging [19], [20] and hyperspectral imaging [21], [22]. Using NMF for hyperspectral "unmixing" — i.e., determining the basic constituent reflectivity spectra ("endmembers") that contribute to each pixel in the image — has been the subject of much research in recent years [23], [24], [25], [26], [27], [28], [29]. The linear mixing model assumed by most hyperspectral unmixing analyses is directly analogous to gamma-ray photon emission, i.e., the measured gamma-ray spectrum is the sum of individual spectra from portions of the environment, and therefore the extension of NMF from hyperspectral imagery to gamma-ray spectroscopy follows naturally.

This work is the first known application of NMF to gamma-ray spectroscopy. To demonstrate the methods described in this work, we consider data collected by the Radiological Multi-sensor Analysis Platform (RadMAP) [30]. In particular, we use data from 99 detectors of the 10×10 array of 10 cm $\times$ 10 cm $\times$ 5 cm NaI detectors recorded at various locations around the San Francisco Bay Area between December 2011 and December 2013. Despite the large array size, binning gamma-ray events into 125-bin histograms at 1 Hz results in spectra with mean rates between 0.9 and 288 counts per bin per second. Due to these low count rates, the Poisson likelihood function is optimized instead of the least squares optimization often used in other fields. The resulting model allows for the interpretation of gamma-ray spectra as being the sum of a small number of non-negative component spectra. When combined with a library containing spectral templates of potential sources of interest, NMF models can be used to simultaneously detect and identify sources by means of statistical tests. Source-injected spectra [31], [32] containing ^{133}Ba and ^{137}Cs are used to quantify performance of the anomaly detection and identification algorithms described in this work.

The remainder of the text is outlined as follows. Section II introduces NMF for background modeling, and source detection and identification. Section III outlines methods used for evaluating the performance of the detection and identification methods, as well as existing methods to compare with, and results are reported in Section IV. Section V concludes with an outlook on additional uses for NMF in gamma-ray spectroscopy.

II. NMF FOR DETECTION AND IDENTIFICATION

A. Non-negative Matrix Factorization

Dimensionality reduction is a procedure for encoding a vector $\mathbf{x} \in \mathbb{R}^d$ as a vector $\mathbf{a} \in \mathbb{R}^k$, where k is typically much less than d, yielding a compact representation of the original data. For linear dimensionality reduction techniques such as PCA, a data matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ consisting of n d-dimensional vectors can be approximated as the product of two low-rank matrices:

$$\hat{\mathbf{X}} = \mathbf{AV} \qquad (1)$$

where $\mathbf{A} \in \mathbb{R}^{n \times k}$ are weights encoding data in the low-dimensional space, and $\mathbf{V} \in \mathbb{R}^{k \times d}$ is a set of k basis vectors spanning a subspace of $\mathbb{R}^d$. Such techniques allow measurements to be approximated as a linear combination of the rows of $\mathbf{V}$, which are components that capture structure in the data being considered. One use of dimensionality reduction is data visualization, in which high-dimensional data $\mathbf{X}$ is projected to two or three dimensions using $k = 2$ or $k = 3$, respectively. The coordinates $\mathbf{A}$ in the low-dimensional subspace can then be visualized, possibly allowing one to see patterns in the data capured by the components in $\mathbf{V}$ that may not have been transparent at the full dimensionality.

NMF [15], [16] is a linear dimensionality reduction technique that approximates a matrix $\mathbf{X}$ in terms of the matrices $\mathbf{A}$ and $\mathbf{V}$, where all three matrices are required to have non-negative entries. Prior to performing NMF, the rank k must be specified, which is in contrast to PCA, where one can vary k without recomputing the decomposition. Performing the factorization in Equation (1) can be seen as a maximum-likelihood estimation (MLE) problem, and given the discrete

nature of the gamma-ray spectra, we consider the Poisson log-likelihood which relates the measured data x_{ij} in $\mathbf{X}$ and the low-rank approximation $\hat{x}_{ij}$ in $\hat{\mathbf{X}}$

$$\ln P(\mathbf{X}|\hat{\mathbf{X}}) = \sum_{i=1}^{n} \sum_{j=1}^{d} -\hat{x}_{ij} + x_{ij} \ln \hat{x}_{ij} - \ln x_{ij}! \quad (2)$$

where the sum is performed over the bins of all spectra in $\mathbf{X}$. In the MLE interpretation, each element $\hat{x}_{ij}$ can then be thought of as the mean rate per energy bin per spectrum.

Maximizing the log-likelihood in equation (2) is equivalent to minimizing the Kullback-Leibler (KL) divergence between $\mathbf{X}$ and $\hat{\mathbf{X}}$. The KL divergence is a quantity from information theory [33] that measures how well a given "true" probability distribution is approximated by an estimate of the distribution. Here, the "true" distribution is given by the measured data $\mathbf{X}$, while the estimate is $\hat{\mathbf{X}}$. Though performing NMF is generally a non-convex problem with no exact solution, meaning solutions tend to be local minima, the error between the data $\mathbf{X}$ and the estimate $\hat{\mathbf{X}} = \mathbf{AV}$, as measured by the KL divergence, has been shown to be monotonically non-increasing under repeated application of the following iterative update rules [34]:

$$a_{il} \leftarrow a_{il} \frac{\sum_j v_{lj} x_{ij}/\hat{x}_{ij}}{\sum_j v_{lj}} \quad (3)$$

$$v_{lj} \leftarrow v_{lj} \frac{\sum_i a_{il} x_{ij}/\hat{x}_{ij}}{\sum_i a_{il}} \quad (4)$$

$$v_{lj} \leftarrow \frac{v_{lj}}{\sum_m v_{lm}} \quad (5)$$

$$a_{il} \leftarrow a_{il} \sum_m v_{lm} \quad (6)$$

Equation (3) is used to update each element a_{il} of $\mathbf{A}$, then equation (4) is used to update each element v_{lj} of $\mathbf{V}$. The rows of $\mathbf{V}$ are then normalized using equation (5), and the weights are updated again using equation (6) based on the normalization of $\mathbf{V}$. We refer to the procedure of using the update rules in equation (3) – equation (6) to perform NMF as Poisson NMF (PNMF).

As a comparison to PNMF, we also consider minimizing a least squares objective function, which we refer to as L2NMF:

$$\mathcal{L}(\mathbf{X}, \hat{\mathbf{X}}) = \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2 \quad (7)$$

where F denotes the Frobenius norm. From an MLE perspective, the objective function in equation (7) implicitly assumes a Gaussian distribution in each bin with mean $\hat{x}_{ij}$ and having equal variance, meaning one could also write it as a log-likelihood maximization, as with the Poisson model in equation (2).

The iterative update rules for L2NMF [34] are given by:

$$
\begin{aligned}
a_{il} &\leftarrow a_{il} \frac{(\mathbf{XV}^\top)_{il}}{(\mathbf{AVV}^\top)_{il}} \\
v_{lj} &\leftarrow v_{lj} \frac{(\mathbf{A}^\top \mathbf{X})_{lj}}{(\mathbf{A}^\top \mathbf{AV})_{lj}} \\
v_{lj} &\leftarrow \frac{v_{lj}}{\sum_m v_{lm}} \\
a_{il} &\leftarrow a_{il} \sum_m v_{lm}
\end{aligned}
\quad (8)
$$

As with the PNMF update rules, the L2NMF update rules are used to sequentially updates weights $\mathbf{A}$, components $\mathbf{V}$, then a normalization of bases is performed.

Given a data matrix $\mathbf{X}$, a desired rank k, and initial values for $\mathbf{A}$ and $\mathbf{V}$ (e.g., using random initialization), equation (3) – equation (6) and equation (8) yield approximate NMF solutions using generalized KL divergence and least squares objective functions, respectively. After initializing $\mathbf{A}$ and $\mathbf{V}$, both matrices are updated in an alternating fashion by applying the multiplicative rules on one while the other is held constant, which we refer to as training. Similar to other iterative methods, the multiplicative updates on $\mathbf{A}$ and $\mathbf{V}$ are typically repeated for a fixed number of steps, or until the value of the reconstruction error, as measured by equation (2) or equation (7), converges within some specified tolerance.

As a result of the iterative nature, NMF is more computationally intensive than other linear dimensionality reduction techniques such as PCA. Additionally, the non-convexity of NMF means that solutions may yield different combinations of weights and basis vectors for different initializations. For the applications described in this work, however, these drawbacks have not been prohibitive. For example, while the training procedure generally requires significantly more time to complete than for PCA, training is typically performed offline, meaning it does not present an operational burden. Given a trained model $\mathbf{V}$, the procedure for encoding a single spectrum with NMF weights is carried out in near real-time, meaning anomaly detection and identification can be performed as spectra are being collected. To address the non-uniqueness of solutions, in this work, each NMF decomposition is initialized using non-negative double singular value decomposition (NNDSVD) [35], which allows for deterministic NMF approximations. NMF solutions generated using NNDSVD-initialized matrices do not necessarily result in optimal performance, with respect to detecting and identifying sources of interest, but these solutions have found to work sufficiently well for the analyses performed in this work.

B. Background Modeling using NMF

In applying NMF to gamma-ray spectroscopy, we describe a spectrum $\mathbf{x} \in \mathbb{R}_+^d$ as a linear combination of k non-negative basis vectors. To perform NMF, n gamma-ray spectra, each having d energy bins and which are representative of the background in the environment to be searched for sources, are used to form a matrix $\mathbf{X} \in \mathbb{R}_+^{n \times d}$. We refer to this process of generating the basis $\mathbf{V}$ on background data as background modeling. For the factorization $\hat{\mathbf{X}} = \mathbf{AV}$, the rows of $\mathbf{V}$ are normalized, such that the weights $\mathbf{A}$ can be interpreted as the number of gamma-ray counts in the spectrum from the corresponding NMF component.

The left side of Fig. 1 shows three components and the associated weights resulting from performing 30,000 PNMF fitting iterations on 86,400 one-second NaI spectra randomly-sampled from various locations around the San Francisco Bay Area. The data used are discussed further in Section III-A. Note that the ordering of resulting NMF components is arbitrary, unlike PCA where lower component numbers capture

more variance in the data. However, the NMF components in Fig. 1 were numbered to match the ordering of the corresponding PCA components shown on the right side of Fig. 1.

The NMF components in Fig. 1 capture features that have physical meaning. Component 1 appears to be the mean spectral shape. Component 2 has a higher magnitude than the other two at energies below 125 keV, which may be indicative of distant sources or multiple scatters. Component 3 is a combination of the ^{40}K 1460 keV line and the ^{208}Tl 2615 keV line, among others, and generally shows more pronounced peaks than Component 1. As a result of the physical interpretation, when encoding a spectrum using NMF, the component weights yield insight to the composition of the gamma-ray background. This interpretation is particularly of interest for the integration of gamma-ray detectors with other environmental sensors, where the latter could potentially provide information about the composition of the local gamma-ray background.

As a comparison with components formed using NMF, the upper right panel of Fig. 1 shows the first three principal components (PCs) found by performing PCA using the correlation matrix generated from $\mathbf{X}$. By comparing the components $\mathbf{V}$ and distributions of weights $\mathbf{A}$ from both NMF and PCA, we see that there are similarities in the structure that is captured within components. The distributions of weights for component 1 for NMF and PCA, which contain the mean spectral shape, are similar both in shape and also in that both capture the most variance. We also see differences in the spectral shapes, the most obvious being that components 2 and 3 from PCA capture both positive and negative spectral features.

C. Anomaly Detection using NMF

Anomaly detection aims to establish whether or not a spectrum is unusual without attempting to identify radiological sources. The NMF formulation described above can be used for anomaly detection by finding the reconstructed low-rank approximation $\hat{\mathbf{x}} = \mathbf{a}^\top \mathbf{V}$ of a spectrum $\mathbf{x}$, and treating $\hat{\mathbf{x}}$ as the mean rate per energy bin per spectrum. Using the mean background rate per energy bin $\hat{x}_i$, the probability of measuring greater than or equal to x_i counts or less than or equal to x_i counts, referred to as the p-value p_i, is calculated for each bin. To account for measurements that are either much higher or lower than the mean rate, we consider the two-tailed p-value

$$p_i = 2\min(F(x_i; \hat{x}_i), 1 - F(x_i; \hat{x}_i)) \tag{9}$$

where $F(x_i; \hat{x}_i)$ is the cumulative density function of the Poisson distribution with mean rate $\hat{x}_i$. We then aggregate the p-values to form the NMF anomaly detection score:

$$\text{AD}_{\{P,L2\}}(\mathbf{x}) = -\frac{1}{d}\sum_{i=1}^{d} \ln p_i. \tag{10}$$

This anomaly detection metric is evaluated over a dataset containing only background spectra, and from this distribution of metrics, we can empirically set a threshold T_Z based on a target FAR, where Z specifies the anomaly detection metric used. When testing a spectrum $\mathbf{x}'$ for the presence

of anomalies, $\text{AD}_Z(\mathbf{x}')$ is compared to T_Z, and an alarm is registered if $\text{AD}_Z(\mathbf{x}')$ exceeds T_Z.

D. Simultaneous Detection and Identification using NMF

With the additive, parts-based interpretation of NMF in mind, the background model $\mathbf{V}$ can be augmented to account for contributions from gamma-ray sources of interest by including a source template $\mathbf{t}_s \in \mathbb{R}_+^d$ corresponding to source s (e.g., ^{133}Ba) with the matrix $\mathbf{V}$ to form a combined background and source model $\mathbf{V}_s = \left[\begin{smallmatrix} \mathbf{V} \\ \mathbf{t}_s^\top \end{smallmatrix}\right]$. By fitting a spectrum to $\mathbf{V}_s$, the spectrum is approximated as a linear combination of the background components and the source. The source template, representing the system's response to a particular source, can be generated from measurements, simulation, or by performing NMF on data containing sources.

The training procedure is used to generate both the components $\mathbf{V}$ and weights $\mathbf{A}$. When examining new spectra $\mathbf{X}'$, the model $\mathbf{V}_s$ is applied and held fixed, resulting in new weights $\mathbf{A}'$. Fig. 2 shows the result of performing an NMF fit to a spectrum using both the background-only model $\mathbf{V}$ and the combined source and background model $\mathbf{V}_s$. The measurement is a one-second NaI spectrum of background data from RadMAP that includes source contribution from an injected 85-μCi ^{133}Ba source at 20m standoff distance. We see that together, the background components generated from NMF and the ^{133}Ba source template perform a better approximation of the spectrum than the background model alone.

To detect and identify sources using the augmented model $\mathbf{V}_s$, we examine a test statistic for determining the presence or absence of a source within a spectrum. A given spectrum $\mathbf{x}$ is fit to both the background model $\mathbf{V}$ and augmented model $\mathbf{V}_s$, yielding the negative log-likelihoods $-\ln P(\mathbf{x}|\mathbf{V})$ and $-\ln P(\mathbf{x}|\mathbf{V}_s)$, respectively, which are then used to form a log-likelihood ratio test statistic

$$D_s(\mathbf{x}) = -2\big\{\ln P(\mathbf{x}|\mathbf{V}) - \ln P(\mathbf{x}|\mathbf{V}_s)\big\}. \tag{11}$$

The test statistic $D_s(\mathbf{x})$ quantifies the amount by which the fit is improved by including the additional source template $\mathbf{t}_s$; if the fit is not significantly enhanced, $D_s(\mathbf{x}) \approx 0$, and $D_s(\mathbf{x})$ increases as including the template improves the fit. The distribution of $D_s(\mathbf{x})$ evaluated on background data free of the source s shows the amount of variation expected in the background, and we can use this information to empirically set a per-template threshold D_s^T based on a desired FAR. Fig. 3 shows histograms of $D_s(\mathbf{x})$ for ^{137}Cs and ^{133}Ba from 55 hours of NaI spectra collected by RadMAP. We see that the distribution peaks at $D_s(\mathbf{x}) = 0$ for both sources, and decreases with increasing $D_s(\mathbf{x})$.

Similar to the procedure described in Section II-C, for a given spectrum $\mathbf{x}'$ to be tested for the presence of source s, $D_s(\mathbf{x}')$ is calculated and compared to the empirically-determined threshold D_s^T. If $D_s(\mathbf{x}')$ exceeds the threshold D_s^T, the algorithm registers an alarm for the source s. The procedure of training models using NMF, fitting spectra using background models and source templates, and performing
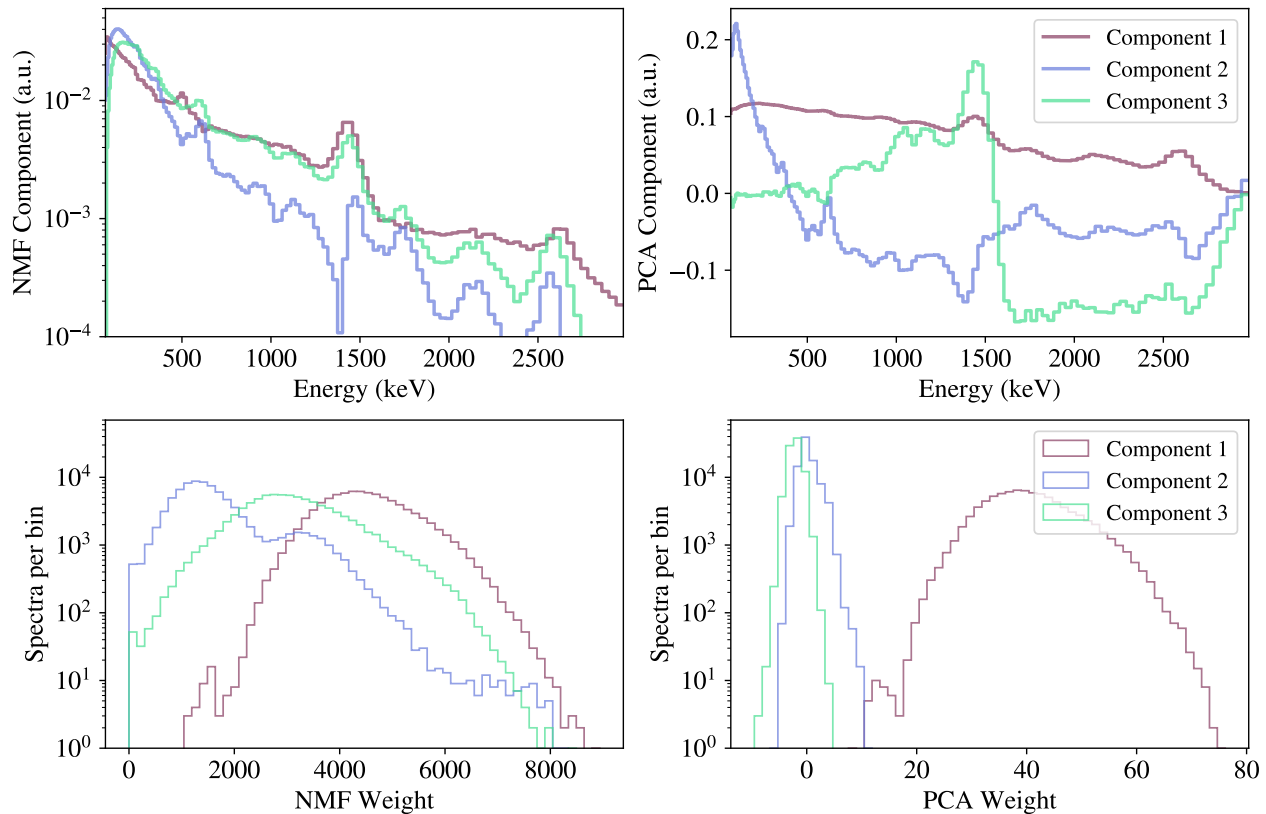
Fig. 1. Three Poisson NMF basis vectors **V** formed from background spectra collected by the RadMAP mobile detection system (upper left) and histograms of weights **A** corresponding to each component (lower left). The first three principal components formed from background spectra (upper right) and histograms of weights corresponding to each component (lower right). For both NMF and PCA, models were formed using 86,400 randomly-sampled spectra, each with one-second integration time, and for NMF, training was done over 30,000 fit iterations. Note that there is no significance in the numbering of NMF components, unlike in PCA, however, NMF components were numbered to match the order of corresponding PCs.

likelihood ratio tests to perform identification is hereafter referred to as ADI_P and ADI_{L2}.

III. BENCHMARKING

A. RadMAP Source Injection Data

Gamma rays with energy between 67 and 3000 keV measured with the NaI array on RadMAP are binned using a one-second integration time into 125-dimensional spectra with bin widths linearly increasing with energy. An approximately 50% opacity coded mask made of lead was on the starboard side of the NaI array during data collection. Note, however, that this mask has no bearing on conclusions made here regarding relative algorithm performance. See [32] for further discussion on the mask and implications to detection sensitivity.

To simulate measuring various source activities, source-injection, in which a source spectrum is downsampled to the desired activity and superimposed on measured background data, is used. Similar to procedures used in previous RadMAP studies [31], [32], segments of data in which the vehicle traveled straight for at least 100 m, within a tolerance of 2.5 m, and maintained a speed of 6.7 ± 1.3 m/s were used as source-injection background data. A set containing 55 hours of data is sampled from for creating background models and threshold setting, and is referred to as the training set. An additional 2 hours of data that met the above criteria are used

as the background for source injections, yielding 315 injected segments, and is referred to as the test set. The background count rates for the spectra used in training models have a mean and standard deviation of 9417 and 1871 counts per second, respectively, while spectra used in source injection have a mean and standard deviation of 9658 and 1942 counts per second, respectively. Fig. 4 shows three randomly-sampled background spectra from the training set, showing some of the variation captured.

Source injections resulting from both background-subtracted unshielded ^{137}Cs measurements and simulations of unshielded ^{133}Ba were then separately added to the aforementioned background segments. P_D is defined as the proportion of segments in which sources are detected or identified. Systematic uncertainties for the source contributions are 11% for the background-subtraction of ^{137}Cs and 10% for simulations of ^{133}Ba. These systematic uncertainties have no effect on the conclusions drawn about relative algorithm performance, and are excluded from tables and figures containing results. Additional details regarding the injection source contributions can be found in [32].

B. Background Model Rank Selection

In performing dimensionality reduction, the rank k of the low-dimensional subspace must be specified. In general, in-
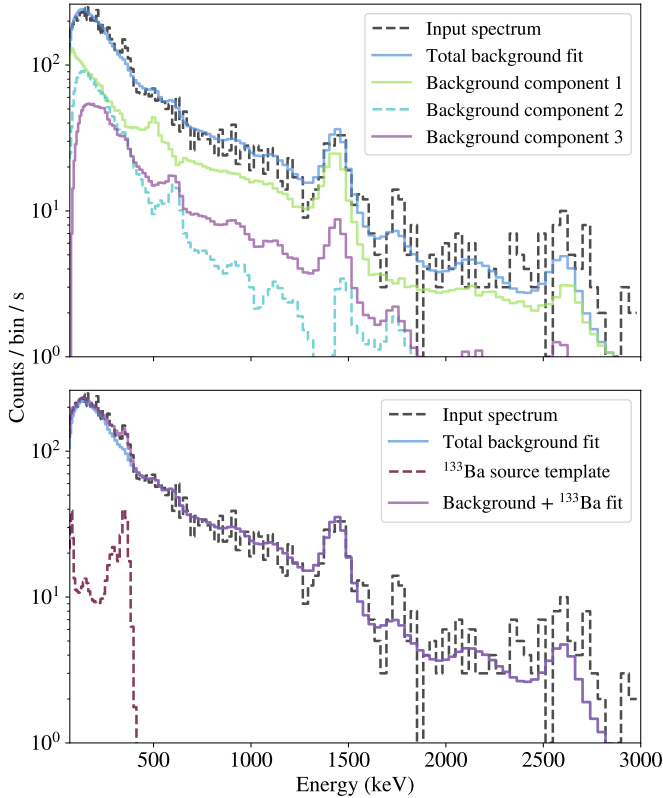
Fig. 2. Comparison of a fit to a spectrum containing ^{133}Ba using only the background model $\mathbf{V}$ (top) and using the background and source model $\mathbf{V}_{\text{Ba-133}}$. The upper pane shows the weighted background components, as well as their sum. The lower pane shows the sum of the background components, the weighted source template, and the sum of the background and source components. By including the source template, the model is able to fit the 356 keV ^{133}Ba peak. An 85-μCi ^{133}Ba source at 20m standoff distance is used.

creasing k will result in enhanced fits. For complete decomposition methods such as PCA and NMF, as $k \to d$, the reconstruction error will vanish, resulting in the loss of ability to detect anomalies. As a result, we aim to choose a value of k that allows us to describe background variations in data sufficiently well, but not at the cost of losing the ability to detect anomalies within spectra.

The Akaike Information Criterion (AIC) [36] is used for determining the number of components used, and is defined as:

$$\text{AIC} = 2\kappa - 2\ln P(\mathbf{X}|\hat{\mathbf{X}}) \tag{12}$$

where $\kappa = kn + kd - k$ is the total number of free parameters in the model $\hat{\mathbf{X}} = \mathbf{AV}$, with k degrees of freedom being removed due to the normalization constraint on $\mathbf{V}$. The value of k that results in the lowest AIC on the training data is used, as it yields the best fit with the fewest parameters.

Equation (12) was used on PNMF and PCA models using $n = 86,400$ 1s spectra, and both yielded a value of $k = 2$ components. AIC for L2NMF was seen to decrease monotonically with k, and as a result $k = 2$ was also used to be consistent with the other models, though additional values of k are also considered for completeness. For readers familiar with PCA, the proportion of variance explained as a function
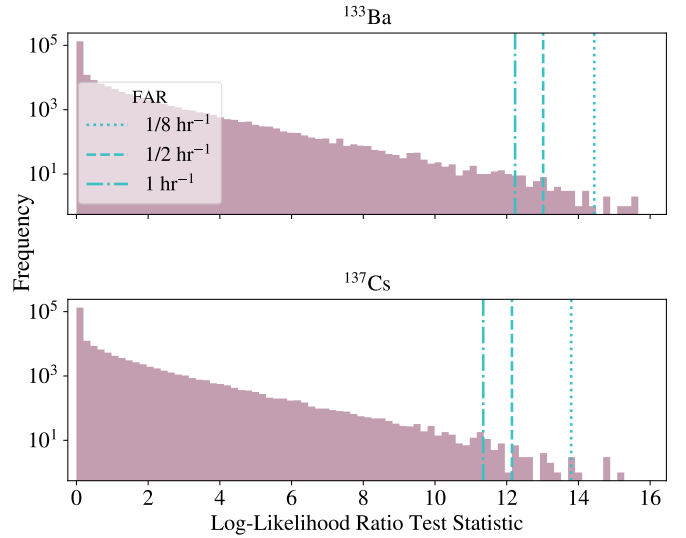

Fig. 3. Distributions of likelihood ratio test statistic $D_s(\mathbf{x})$ used in testing for the presence of ^{133}Ba (top) and ^{137}Cs (bottom). The distributions are formed by computing the difference of negative log-likelihoods between source and background models $\mathbf{V}_s$ and background only models $\mathbf{V}$ over a background dataset containing 55 hours of NaI spectra collected by RadMAP. The distribution is used to empirically select a threshold based on a target FAR. The vertical lines show thresholds for particular FARs, given in the legend.
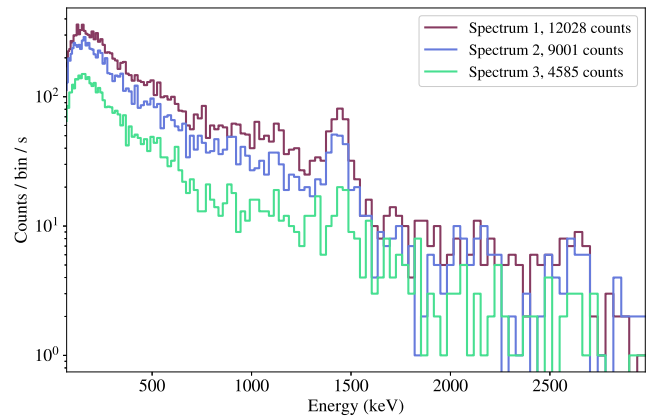

Fig. 4. Three randomly-sampled background spectra from the training set, each labeled with the respective number of total counts in the spectrum. From these three spectra, we can see that samples vary not only by magnitude, but also by spectral features. Spectral variability is captured by the components in the NMF and PCA models described in this work, as well as the empirical threshold setting using for anomaly detection and identification algorithms. Note that such diverse spectra are also present in the background spectra used for source injection.

of k was also examined for PCA and suggested a value of $k = 2$, in agreement with the AIC. Each PNMF, L2NMF, and PCA model described in the following sections will then use $k = 2$ components trained over a single subset of 24 hours of background data sampled from the 55 hour training set. Each model is then evaluated over the entire 55 hour training set in empirically determining thresholds.

C. AD_l and ADI_l in the Limit of Known Background

To understand the limitations of the NMF-based anomaly detection and identification algorithms, we consider the performance of each algorithm in the ideal case, where the mean background rate for each spectrum is known. That is, for each set of injected spectra $\mathbf{X}_{\text{test}}$ consisting of background and source contributions (i.e., $\mathbf{X}_{\text{test}} = \mathbf{X}_{\text{bkg}} + \mathbf{X}_{\text{src}}$), we assume that the mean background rate $\hat{\mathbf{X}}_{\text{bkg}}$ resulting from performing NMF on each spectrum is known. Note that the background used here $\mathbf{X}_{\text{bkg}}$ are the 2 hours of background data referred to previously, which are separate measurements from the training data used in generating models.

In this best case scenario, algorithm performance is primarily limited by statistical fluctuations in the background term $\mathbf{X}_{\text{bkg}}$, and we refer to this as the known background limit (KBL). One reason for exploring the KBL is that this procedure lends insight to how much performance can be improved with an enhanced understanding of the local radiological background.

To determine KBL performance, we use the following procedure:

1) Project the measured background contribution $\mathbf{X}_{\text{bkg}}$ onto the subspace generated by $\mathbf{V}$ to yield the NMF weights $\mathbf{A}_{\text{bkg}}$.
2) Estimate the mean background rate as $\hat{\mathbf{X}}_{\text{bkg}} = \mathbf{A}_{\text{bkg}}\mathbf{V}$.
3) To detect anomalies, use $\hat{\mathbf{X}}_{\text{bkg}}$ as the reconstructed spectra to which the measured spectra $\mathbf{X}$ are compared.
4) To perform identification, hold the background contribution $\hat{\mathbf{X}}_{\text{bkg}}$ fixed, and find the source contribution that best fits the source and background model to the spectra $\mathbf{X}$.

For both anomaly detection and simultaneous detection and identification, thresholds are computed as previously described, and metrics from test spectra are compared to these thresholds to determine the presence or absence of sources.

D. Anomaly Detection Benchmarking

In section IV-A two existing methods, gross counts $K\sigma$, referred to here simply as $K\sigma$, and PCA-based anomaly detection, are compared to AD_P and AD_{L2}. $K\sigma$ measures how many standard deviations σ the gross counts in a spectrum are away from the mean number of background counts μ as:

$$K = \frac{\|\mathbf{x}\|_1 - \mu}{\sigma}, \tag{13}$$

where the 1 subscript in the denominator denotes the L1 norm of $\mathbf{x}$, or gross counts. Here, the mean μ and standard deviation σ of the total number of counts within a spectrum are determined from the training set.

A PCA-based alarm metric [37], [12], [38] uses the first k PCs $\tilde{\mathbf{V}}$ found from PCA to measure deviations between a spectrum and its low-dimensional PCA representation

$$\text{AD}_{PC}(\mathbf{x}) = \frac{\|(\mathbb{I} - \tilde{\mathbf{V}}^{\top}\tilde{\mathbf{V}})\mathbf{x}\|_2}{\sqrt{\|\mathbf{x}\|_1}}. \tag{14}$$

This normalization is included to account for larger residuals at higher count rates, which is in contrast to the p-value based

TABLE I
REGION OF INTEREST ENERGY WINDOWS (KEV)

Source	$\mathcal{R}$	$\mathcal{B}_1$	$\mathcal{B}_2$
^{133}Ba	340.0 - 397.2	397.2 - 443.0	443.0 - 474.9
^{137}Cs	632.3 - 670.4	670.4 - 749.8	749.8 - 812.2

anomaly detection metric that accounts for variation by means of a Poisson model for each energy bin. As opposed to standard PCA, which computes the singular value decomposition (SVD) on the covariance matrix, the basis $\tilde{\mathbf{V}}$ is formed by performing the SVD on the correlation matrix.

E. Region of Interest (ROI)

To assess the relative performance of the ADI_P and ADI_{L2} algorithms, a Region of Interest (ROI) algorithm was evaluated on the same data. The ROI algorithm, as described in [32], estimates the number of source counts within a spectral region in which gamma rays from a particular source are expected. To estimate source counts in the ROI $\mathcal{R}$, the algorithm first uses a linear relationship between the number of background counts in the ROI and the number of counts in neighboring regions $\mathcal{B}_1$ and $\mathcal{B}_2$. Regions $\mathcal{B}_1$ and $\mathcal{B}_2$ are chosen to be at higher energies than R so that they do not contain counts from downscattered source gamma rays, while being sufficiently close to $\mathcal{R}$ to predict background counts in $\mathcal{R}$. The window edges defining the ROI and the background regions are shown in Table I.

The number of background counts in $\mathcal{R}$ is estimated as $\hat{r} = \mathbf{b}^{\top}\mathbf{w}$, where $\mathbf{b}$ is a vector containing counts within the background windows $\mathcal{B}_1$ and $\mathcal{B}_2$, and $\mathbf{w}$ are weights found via Poisson regression [39] on the same background training set used for training NMF models. When evaluating spectra for the presence of the source, the difference between the measured counts r and estimated background counts $\hat{r}$, or the residual, forms the ROI metric, defined as

$$\text{ROI}(\mathbf{x}) = \frac{r - \hat{r}}{\sqrt{\hat{r}}}. \tag{15}$$

This metric is then compared to a decision threshold to determine the presence or absence of a source. As with the other methods described in this work, the decision thresholds are computed empirically for a given FAR by calculating the residual over background data. Spectra in which the ROI has an excess of counts are then considered anomalous.

F. Probability of Detection

Similar to [32], the performance of the ADI_P and ADI_{L2} algorithms is evaluated by calculating the probability of detection P_D as a function of source activity. In particular, we consider source drive-bys in which RadMAP passes a source of a specific activity within a background environment. A total of 315 source drive-bys are performed, and P_D is empirically determined by the proportion of drive-bys in which the source was detected. Empirical values of P_D are then fit to a sigmoid function of the form

$$P_D(x) = \frac{1}{1 + e^{-(x-\mu)/\omega}} \tag{16}$$

Using this relationship, we determine the minimum detectable activity at 95% probability of detection (MDA$_{95}$) at a fixed FAR. The Jeffreys interval [40] is used to determine the 68% confidence interval on the probability of detection.

IV. Results

A. Anomaly Detection

As with the procedure for simultaneous detection and identification, values for NMF-based and PCA-based anomaly detection scores were calculated over the background training set **X**, and a threshold on scores were empirically determined using a 1/8 hr^{-1} (3.5×10^{-5}s^{-1}) FAR. For anomaly detection, however, only a single threshold is determined, whereas identification requires a threshold for each source. The same 24 hour sample of spectra used in training the NMF models was used in training PCA models, and the entire 55 hour set of data was used for threshold setting.

Fig. 5 shows a comparison of probability of detection curves for NMF and PCA-based anomaly detection algorithms, each using $k = 2$. Table II shows the corresponding MDA$_{95}$ resulting from a fit of P_D to a sigmoid for the $k = 2$ models, as well as models using $k = 4$. Note that models with $k > 4$ were also considered, but performance did not improve significantly, and results are omitted. From Fig. 5 and Table II, we see that AD$_P$ and AD$_{L2}$ are able to maintain an MDA near 200 μCi for both sources, while AD$_{PC}$ has an increased MDA for ^{137}Cs, which is likely resulting from the difference in anomaly detection metrics between the NMF and PCA-based models. Note, however, that for both sources and for both $k = 2$ and $k = 4$, AD$_P$ outperforms AD$_{L2}$, which is perhaps attributed to the use of a more accurate statistical model in AD$_P$. Furthermore, in detecting ^{137}Cs, AD$_P$ has a similar MDA to AD$_{PC}$ KBL. That is, by using a more accurate statistical model and computing a different anomaly detection score, AD$_P$ is able to detect weaker sources than the PCA-based method under a near-perfect understanding of the gamma-ray background. However, AD$_{PC}$ offers a slight advantage over AD$_{L2}$ for ^{133}Ba, suggesting that these performance enhancements may primarily exist for sources with gamma-rays at higher energies.

Not shown in Fig. 5 are curves for $K\sigma$. As Table II shows, the MDA of $K\sigma$ is much higher than any of the NMF and PCA-based methods. The poor performance of $K\sigma$ reflects the background variability seen by RadMAP, as high thresholds must be set to maintain the FAR of 1/8 hr^{-1}.

When comparing the performance of the dimensionality reduction-based anomaly detection methods with their respective KBLs, we see that there is room for improvement. By including additional information about the local gamma-ray environment, it is possible that these methods approach the case of the known background. However, the NMF-based methods are better suited for coupling with environmental sensors, due to the physical interpretation of NMF.

In using empirical thresholds, results are ultimately dependent on the data used to estimate thresholds. To understand the effect that the choice of dataset used for determining thresholds has on performance, we consider an approach which
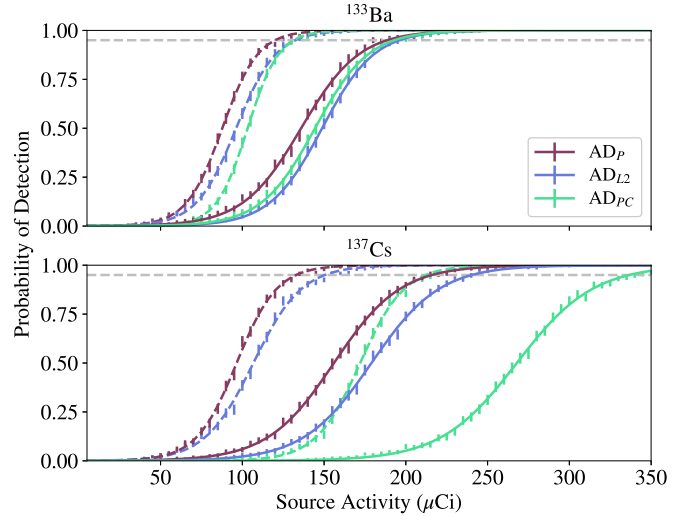

Fig. 5. Probability of detection of ^{133}Ba (top) and ^{137}Cs (bottom) sources at 20m standoff using NMF-based detection algorithms, a PCA-based detection algorithm, and both NMF and PCA-based algorithms in the case of a known background, each model using $k = 2$ components at 1/8 hr $^{-1}$ FAR and 1s integration time. Solid lines show the measured performance, and dashed lines show performance in the KBL. Note that only anomaly detection is performed here, meaning sources are not identified by the algorithms. The 95% probability of detection is indicated by the dashed horizontal line. Error bars indicate the 68% Jeffreys interval. Not shown is a curve for gross counts, which has a significantly higher MDA.

TABLE II
MDA$_{95}$ (μCI + STATISTICAL ERROR) FOR ANOMALY DETECTION

Source / Method	^{133}Ba	^{137}Cs
AD$_P$ KBL, $k = 2$	119.9 ± 0.3	132.7 ± 0.4
AD$_{L2}$ KBL, $k = 2$	132.2 ± 0.3	150.3 ± 0.6
AD$_{PC}$ KBL, $k = 2$	131.2 ± 0.2	212.0 ± 0.5
AD$_P$, $k = 2$	188.5 ± 0.7	215.9 ± 0.5
AD$_P$, $k = 4$	181.3 ± 0.9	209.4 ± 0.7
AD$_{L2}$, $k = 2$	195.9 ± 0.4	240.3 ± 0.6
AD$_{L2}$, $k = 4$	182.4 ± 0.5	231.1 ± 0.8
AD$_{PC}$, $k = 2$	192.9 ± 0.4	336.5 ± 0.7
AD$_{PC}$, $k = 4$	189.1 ± 0.5	325.8 ± 0.9
$K\sigma$	1971.0 ± 13.6	2302.2 ± 16.1

estimates thresholds based on random subsets of data. In particular, the data is split into many contiguous segments of measurements, and a number of these segments are picked at random, without replacement, and used for computing thresholds. Such random sampling was performed 100 times, with each iteration having a combined measurement time of approximately 24h. Thresholds for AD$_P$, AD$_{L2}$, and AD$_{PC}$ were computed for each of the 100 trials, resulting in a distribution of thresholds for each anomaly detection method. The estimated mean and standard deviation of each metric, along with thresholds generated using the entire 55h dataset, are given for $k = 2$ models in Table III. This table shows that thresholds estimated using the entire dataset fall within one standard deviation of the mean of threshold generated using random subsets. The mean and standard deviation of

TABLE III
COMPARISON OF EMPIRICAL ANOMALY DETECTION THRESHOLDS

Threshold	$\mathbf{AD}_P$	$\mathbf{AD}_{L2}$	$\mathbf{AD}_{PC}$
Full dataset	1.337	1.420	1.611
Random subsets	1.341 ± 0.007	1.415 ± 0.012	1.613 ± 0.017

TABLE IV
LOWER AND UPPER ESTIMATES OF MDA_{95} (μCi + STATISTICAL ERROR)
FOR ANOMALY DETECTION

Method \ Source	^{133}Ba	^{137}Cs
AD_P, $k = 2$	$(187.5 \pm 0.6,$ $191.9 \pm 0.6)$	$(214.5 \pm 0.5,$ $220.8 \pm 0.5)$
AD_{L2}, $k = 2$	$(194.5 \pm 0.4,$ $199.9 \pm 0.4)$	$(241.1 \pm 0.7,$ $251.9 \pm 0.8)$
AD_{PC}, $k = 2$	$(190.4 \pm 0.4,$ $196.3 \pm 0.5)$	$(331.4 \pm 0.7,$ $343.4 \pm 0.7)$

these thresholds are then used to generate a lower and upper estimates for thresholds, yielding a lower and upper estimate on MDA. Lower and upper values for MDA using thresholds equal to a standard deviation below and above the mean, respectively, are reported in Table IV. The range of MDA values presented in Table IV suggest that the procedure for computing AD_P may result in empirical thresholds that are less sensitive to data than AD_{L2} and AD_{PC}.

B. Simultaneous Detection and Identification

Using a FAR of $1/8$ hr^{-1}, thresholds for both sources were empirically determined, yielding the values $D_{\text{Cs}}^T = 14.1$ and $D_{\text{Ba}}^T = 14.4$. Additionally, an L2NMF model was formed and thresholds were calculated using the same training set as the PNMF model described above, though we omit thresholds and figures specific to L2NMF for the sake of brevity. Note that in each case, we are only searching for a single source. In the case of multiple sources being separately tested for, false alarm rates should be adjusted accordingly (e.g., using the Bonferroni correction), ultimately increasing thresholds and reducing sensitivity.

Fig. 6 shows the probability of detection for ^{133}Ba and ^{137}Cs using both NMF-based algorithms and an ROI algorithm, and Table V shows the corresponding MDA_{95} for each algorithm resulting from a fit of P_D to a sigmoid. From Fig. 6, we can see that there is a significant improvement in detection capabilities by using NMF-based algorithms, and that there are additional improvements by using PNMF models over L2NMF.

Similar to results shown in IV-A, there is a difference between measured performance of the NMF-based detection and identification methods and the performance in the case of a known background. By informing these methods with data about the local environment, more accurate background models could potentially be created, enhancing detection and identification performance.
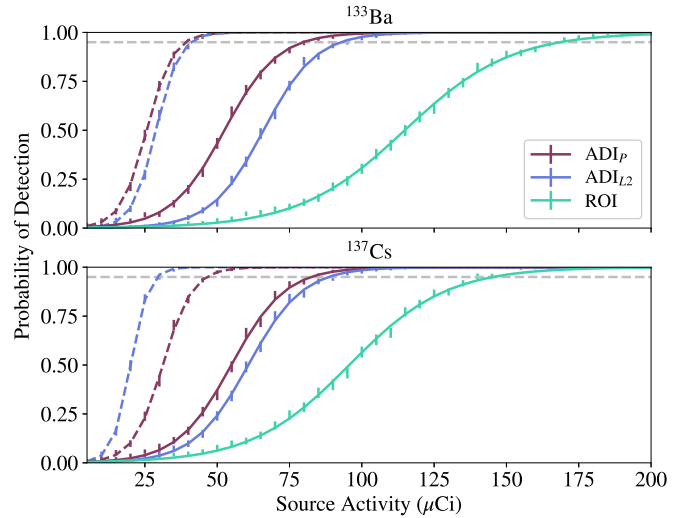

Fig. 6. Probability of detection and identification for ^{133}Ba (top) and ^{137}Cs (bottom) using the ADI_P, ADI_{L2}, NMF-based algorithms in the case of a known background, each using, $k = 2$ components, and an ROI algorithm. Solid lines show the measured performance, and dashed lines show performance in the KBL. Spectra were formed using 1s integration time for 99 NaI detectors on RadMAP, and source injection was performed using 20m standoff distance and 6.7 ± 1.3 m/s vehicle speed. For both sources, a 1/8 hr^{-1} FAR was used. Both the ADI_P and ADI_{L2} methods are able to achieve a 95% probability of detection, indicated by the dashed horizontal line, at a much lower activity than the ROI algorithm. Error bars indicate the 68% Jeffreys interval.

TABLE V
MDA_{95} (μCi + STATISTICAL ERROR) FOR DETECTION AND
IDENTIFICATION

Method \ Source	^{133}Ba	^{137}Cs
ADI_P KBL, $k = 2$	38.9 ± 0.1	46.3 ± 0.2
ADI_{L2} KBL, $k = 2$	40.8 ± 0.1	29.1 ± 0.1
ADI_P, $k = 2$	79.6 ± 0.3	82.4 ± 0.4
ADI_{L2}, $k = 2$	92.7 ± 0.3	88.4 ± 0.3
ROI	168.5 ± 0.6	145.2 ± 0.8

V. DISCUSSION

Detecting and identifying gamma-ray sources relies on the ability of an algorithm to distinguish source and background. Here, we proposed of the use of NMF, under the assumption of Poisson statistics, to model radiological backgrounds. These models can be used to find radiological anomalies, and by using spectral templates, source identification can be performed. In addition to performing identification for single sources, multiple sources can be simultaneously tested for by appending additional source templates to the matrix $\mathbf{V}_s$. Under the given parameters, the anomaly detection and identification algorithms outperform standard algorithms. Note, however, that the decomposition-based methods here were only optimized with respect to the number of degrees of freedom, and that the performance of each method, including the benchmark algorithms, may be enhanced by optimizing integration time, performing spectral smoothing, normalizing input variables, and optimizing energy windows in the case of ROI.

While the methods in this work were demonstrated using a

large and unique detector system, these methods are general enough to be deployed on a wide range of systems. Furthermore, these methods can be extended to detect and identify gamma-ray sources other than the two studied in this analysis, albeit with additional care. One consideration is the similarity in spectral shape between the source of interest and the NMF components used in estimating background – sources with a high degree of similarity to background components may be incorrectly attributed to background, ultimately reducing the ability to detect the particular source. If a source is seen to exhibit such behavior, one may devise a means of reducing the similarity, for example leveraging the non-uniqueness and selecting background components that are the most distinct from the sources of interest (e.g., with respect to cosine similarity or KL divergence). Similarly, when searching for multiple sources, one must be aware of the similarity between sources, as source templates that contain overlapping gamma-ray lines may result in misidentification. Lastly, though only unshielded sources were considered in this work, which admittedly is a best case scenario, the methods described in this work can be extended to use shielded templates as well.

Beyond anomaly detection and identification, Poisson NMF can be considered a general framework for approaching gamma-ray spectroscopic analyses, as it lends itself to a useful physical interpretation due to its additive, non-negative nature. For example, the geospatial distribution of NMF weights may yield the environmental composition of particular radioisotopes, which could potentially find use in applications such as contamination mapping.

From Fig. 5 and Fig. 6, we see that there is a significant difference between the measured performance and the performance in the limit of known background. Additional information about the environment, for example, in the form of Bayesian priors and regularization, may potentially enhance detection and identification performance. Prior probabilities of background weights $\mathbf{A}$, determined from previous measurements, could be introduced to provide constraints to background models by means of maximum a posteriori, as opposed to the MLE used here, which may allow for a more accurate background estimate. Additionally, performance can be enhanced further by aggregating test statistics over several spectra, or by sequentially estimating weights (e.g., using a Kalman filter) to create time-dependent models.

References

[1] D. M. Pfund, R. C. Runkle, K. K. Anderson, and K. D. Jarman, "Examination of Count-starved Gamma Spectra Using the Method of Spectral Comparison Ratios," *IEEE Transactions on Nuclear Science*, vol. 54, no. 4, pp. 1232–1238, Aug 2007.

[2] C. J. Sullivan, M. E. Martinez, and S. E. Garner, "Wavelet Analysis of Sodium Iodide Spectra," *IEEE Transactions on Nuclear Science*, vol. 53, no. 5, pp. 2916–2922, Oct 2006.

[3] D. Boardman, M. Reinhard, and A. Flynn, "Principal Component Analysis of Gamma-Ray Spectra for Radiation Portal Monitors," *IEEE Transactions on Nuclear Science*, vol. 59, no. 1, pp. 154–160, Feb 2012.

[4] B. R. Cosofret, K. N. Shokhirev, P. A. Mulhall, D. Payne, B. Harris, E. Arsenault, and R. Moro, "Utilization of advanced clutter suppression algorithms for improved spectroscopic portal capability against radionuclide threats," in *2013 IEEE International Conference on Technologies for Homeland Security (HST)*, Nov 2013, pp. 618–622.

[5] D. K. Fagan, S. M. Robinson, and R. C. Runkle, "Statistical methods applied to gamma-ray spectroscopy algorithms in nuclear security missions," *Applied Radiation and Isotopes*, vol. 70, no. 10, pp. 2428 – 2439, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0969804312003818

[6] S. E. Labov, L. Pleasance, P. Sokkappa, W. Craig, G. Chapline, M. Frank, J. Gronberg, J. G. Jernigan, S. Johnson, J. Kammeraad, D. Lange, A. Meyer, K. Nelson, B. Pohl, D. Wright, and R. Wurtz, "Foundations for Improvements to Passive Detection Systems - Final Report," 2004.

[7] K. D. Jarman, R. C. Runkle, K. K. Anderson, and D. M. Pfund, "A comparison of simple algorithms for gamma-ray spectrometers in radioactive source search applications," *Applied Radiation and Isotopes*, vol. 66, no. 3, pp. 362 – 371, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0969804307002886

[8] J. Hovgaard and R. Grasty, "Reducing Statistical Noise in Airborne Gamma-Ray Data Through Spectral Component Analysis," in *Proceedings of Exploration 97*, 1997, pp. 753–764.

[9] B. Minty and P. McFadden, "Improved NASVD smoothing of airborne gamma-ray spectra," *Exploration Geophysics*, vol. 29, no. 3/4, pp. 516–523, Dec. 1998. [Online]. Available: http://library.seg.org/doi/abs/10.1071/EG998516

[10] H. K. Aage, U. Korsbech, K. Bargholz, and J. Hovgaard, "A new technique for processing airborne gamma ray spectrometry data for mapping low level contaminations," *Applied Radiation and Isotopes*, vol. 51, no. 6, pp. 651–662, Dec. 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0969804399000871

[11] J. A. Kulisek, J. E. Schweppe, S. C. Stave, B. E. Bernacki, D. V. Jordan, T. N. Stewart, C. E. Seifert, and W. J. Kernan, "Real-time airborne gamma-ray background estimation using NASVD with MLE and radiation transport for calibration," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 784, pp. 287–292, Jun. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900214014247

[12] P. Tandon, "Bayesian Aggregation of Evidence for Detection and Characterization of Patterns in Multiple Noisy Observations," Ph.D., Carnegie Mellon University, United States – Pennsylvania, 2015. [Online]. Available: https://search.proquest.com/docview/1774007802/abstract/463397BF9CAF4561PQ/1

[13] M. Collins, S. Dasgupta, and R. E. Schapire, "A generalization of principal component analysis to the exponential family," in *Advances in Neural Information Processing Systems*. MIT Press, 2001.

[14] P. Tandon, P. Huggins, A. Dubrawski, S. Labov, and K. Nelson, "Suppressing Background Radiation Using Poisson Principal Component Analysis," *arXiv:1605.08455 [physics, stat]*, May 2016, arXiv: 1605.08455. [Online]. Available: http://arxiv.org/abs/1605.08455

[15] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, Jun. 1994. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/env.3170050203

[16] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788 EP –, 10 1999. [Online]. Available: http://dx.doi.org/10.1038/44565

[17] S. Z. Li, X. W. Hou, H. J. Zhang, and Q. S. Cheng, "Learning spatially localized, parts-based representation," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 2001, pp. I–207–I–212 vol.1.

[18] E. Lee, C. K. Chan, and P. Paatero, "Application of positive matrix factorization in source apportionment of particulate pollutants in Hong Kong," *Atmospheric Environment*, vol. 33, no. 19, pp. 3201–3212, Aug. 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1352231099001132

[19] P. Sajda, S. Du, and L. C. Parra, "Recovery of constituent spectra using non-negative matrix factorization," in *Wavelets: Applications in Signal and Image Processing X*, vol. 5207. International Society for Optics and Photonics, Nov. 2003, pp. 321–332. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/5207/0000/Recovery-of-constituent-spectra-using-non-negative-matrix-factorization/10.1117/12.504676.short

[20] P. Sajda, S. Du, T. R. Brown, R. Stoyanova, D. C. Shungu, X. Mao, and L. C. Parra, "Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain," *IEEE Transactions on Medical Imaging*, vol. 23, no. 12, pp. 1453–1465, Dec. 2004.

[21] C.-Y. Liou and K.-D. Ou Yang, "Unsupervised classification of remote sensing imagery with non-negative matrix factorization," in *ICONIP*, 2005.

[22] V. P. Pauca, J. Piper, and R. J. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra and its Applications*, vol. 416, no. 1, pp. 29–47, Jul. 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S002437950500340X

[23] L. Miao and H. Qi, "Endmember Extraction From Highly Mixed Data Using Minimum Volume Constrained Nonnegative Matrix Factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[24] S. Jia and Y. Qian, "Constrained Nonnegative Matrix Factorization for Hyperspectral Unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 161–173, Jan. 2009.

[25] A. Huck, M. Guillaume, and J. Blanc-Talon, "Minimum Dispersion Constrained Nonnegative Matrix Factorization to Unmix Hyperspectral Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2590–2602, Jun. 2010.

[26] X. Liu, W. Xia, B. Wang, and L. Zhang, "An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 2, pp. 757–772, Feb. 2011.

[27] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral Unmixing via $_{1/2}$ Sparsity-Constrained Nonnegative Matrix Factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.

[28] N. Wang, B. Du, and L. Zhang, "An Endmember Dissimilarity Constrained Non-Negative Matrix Factorization Method for Hyperspectral Unmixing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 554–569, Apr. 2013.

[29] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold Regularized Sparse NMF for Hyperspectral Unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2815–2826, May 2013.

[30] M. S. Bandstra, T. J. Aucott, E. Brubaker, D. H. Chivers, R. J. Cooper, J. C. Curtis, J. R. Davis, T. H. Joshi, J. Kua, R. Meyer, V. Negut, M. Quinlan, B. J. Quiter, S. Srinivasan, A. Zakhor, R. Zhang, and K. Vetter, "RadMAP: The Radiological Multi-sensor Analysis Platform," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 840, pp. 59–68, Dec. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900216309780

[31] T. J. Aucott, M. S. Bandstra, V. Negut, J. C. Curtis, D. H. Chivers, and K. Vetter, "Effects of Background on Gamma-Ray Detection for Mobile Spectroscopy and Imaging Systems," *IEEE Transactions on Nuclear Science*, vol. 61, no. 2, pp. 985–991, April 2014.

[32] T. H. Joshi, R. J. Cooper, J. Curtis, M. Bandstra, B. R. Cosofret, K. Shokhirev, and D. Konno, "A Comparison of the Detection Sensitivity of the Poisson Clutter Split and Region of Interest Algorithms on the RadMAP Mobile System," *IEEE Transactions on Nuclear Science*, vol. 63, no. 2, pp. 1218–1226, April 2016.

[33] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 03 1951. [Online]. Available: https://doi.org/10.1214/aoms/1177729694

[34] D. D. Lee and H. S. Seung, "Algorithms for Non-negative Matrix Factorization," in *Proceedings of the 13th International Conference on Neural Information Processing Systems*, ser. NIPS'00. Cambridge, MA, USA: MIT Press, 2000, pp. 535–541. [Online]. Available: http://dl.acm.org/citation.cfm?id=3008751.3008829

[35] C. Boutsidis and E. Gallopoulos, "Svd based initialization: A head start for nonnegative matrix factorization," *Pattern Recognition*, vol. 41, no. 4, pp. 1350 – 1362, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320307004359

[36] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, December 1974.

[37] K. Miller and A. Dubrawski, "Gamma-ray source detection with small sensors," *IEEE Transactions on Nuclear Science*, vol. 65, no. 4, pp. 1047–1058, April 2018.

[38] P. Tandon, P. Huggins, R. Maclachlan, A. Dubrawski, K. Nelson, and S. Labov, "Detection of radioactive sources in urban scenes using bayesian aggregation of data from mobile spectrometers," *Information Systems*, vol. 57, pp. 195 – 206, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306437915001866

[39] J. A. Nelder and R. W. M. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society. Series A (General)*, vol. 135, no. 3, pp. 370–384, 1972. [Online]. Available: http://www.jstor.org/stable/2344614

[40] L. D. Brown, T. T. Cai, and A. DasGupta, "Interval estimation for a binomial proportion," *Statistical Science*, vol. 16, no. 2, pp. 128–133, 2001. [Online]. Available: http://www.jstor.org/stable/2676790