Radiation Measurements 46 (2011) 1138–1146

Contents lists available at SciVerse ScienceDirect

Radiation Measurements

journal homepage: www.elsevier.com/locate/radmeas

Detection of radionuclides from weak and poorly resolved spectra using Lasso and subsampling techniques

Er-Wei Bai[a,*], Kung-sik Chan[b], William Eichinger[c], Paul Kump[a]

[a] Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242, USA
[b] Department of Statistical and Actuarial Science, University of Iowa, Iowa City, IA 52242, USA
[c] Department of Civil and Environmental Engineering, University of Iowa, Iowa City, IA 52242, USA

ARTICLE INFO

Article history:
Received 29 April 2011
Received in revised form
21 August 2011
Accepted 22 August 2011

Keywords:
Lasso
Subsampling
Radionuclide
Detection
Identification

ABSTRACT

We consider a problem of identification of nuclides from weak and poorly resolved spectra. A two stage algorithm is proposed and tested based on the principle of majority voting. The idea is to model gamma-ray counts as Poisson processes. Then, the average part is taken to be the model and the difference between the observed gamma-ray counts and the average is considered as random noise. In the linear part, the unknown coefficients correspond to if isotopes of interest are present or absent. Lasso types of algorithms are applied to find non-vanishing coefficients. Since Lasso or any prediction error based algorithm is inconsistent with variable selection for finite data length, an estimate of parameter distribution based on subsampling techniques is added in addition to Lasso. Simulation examples are provided in which the traditional peak detection algorithms fail to work and the proposed two stage algorithm performs well in terms of both the False Negative and False Positive errors.

1. Introduction

There has long been an interest in detecting the covert movement of radioactive materials. With the collapse of the Soviet Union, the poor security associated with its nuclear stockpile and the War on Terror, there is an increased interest in uncovering smuggled nuclear materials. This has resulted in reassessment of the potential nuclear signatures of these materials which could be used to detect them, particularly with respect to the emission of gamma-rays. While a number of active detection methods have been proposed (Fetter et al., 1990; Moss et al., 2002; Weller et al., 2006) which can detect relatively small amounts of material, they are ultimately not suitable for use in public areas because they require potentially lethal doses of some type of radiation to be delivered to the object to be examined. This leaves passive sensing as the primary candidate method for use in areas where human is present. A number of both portable and non-portable gamma-ray spectrometers are commercially available; all suitable for such applications as checkpoint monitoring and nuclear searches where the detector and potential source are close and the expected signal is relatively large so that the gamma-ray signature peaks are well-resolved (August and Whitlock, 2005). Gamma-ray signature recognition is straightforward if signals are strong and high-resolution detectors can be used. Such detectors allow unambiguous identification of radioactive nuclide using photo peak search algorithms (Murray, 1998). However, there exist a large number of applications in which the detectors cannot reasonably be expected to be close to the sources, the exposure time may be short and there may have a wide range of materials between the source and the detector. In these cases, the detectors generally used have large volume (to increase sensitivity) and poor resolution. The signals from these detectors will be weak and difficult to separate from the background radiation or from the signatures of commonly used radioactive materials. This paper addresses the problem of identification of nuclides from weak and poorly resolved spectra. Despite the fact that we know in detail the gamma-ray emission spectra of the individual radioactive nuclide of interest, the exact nature of the gamma-ray signatures from the actual objects of interest are unknown. Signals from commonly encountered natural and man-made gamma-ray sources can compete with the signals of interest and cause false alarms. The emissions from nuclear weapons materials Pu239 for example are quite weak, with many of the lines at relatively low energies so that they are easily attenuated. In spite of these difficulties, nuclear weapons materials spectra still remain unique, but may be subtle and difficult to recognize by traditional peak searching detection algorithms.

* Corresponding author. Tel.: +1 319 335 5949.
E-mail addresses: er-wei-bai@uiowa.edu (E.-W. Bai), kung-sik-chan@uiowa.edu (K.-s. Chan), william-eichinger@uiowa.edu (W. Eichinger).

1350-4487/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.radmeas.2011.08.020

Another existing approach is transforming the measurements into a vector of background-adjusted difference referred to as the spectral comparison ratio (SCR) method (Wei et al., 2010; Anderson et al., 2008). Properties of the ratios can be used to determine whether the unknown measurement is similar to that of the benign measurement.

The contribution of the paper is the development of a new two stage detection algorithm that combines recent regression methodologies with subsampling techniques based on the idea of majority voting. The algorithm is built on a Poisson model of gamma-ray emission that leads to a linear regression model. Then, an algorithm referred to as the Lasso together with subsampling techniques are applied. Theoretical and simulation studies reported in this paper demonstrate promising results for the proposed algorithm.

2. Modeling

The emission of gamma-ray from a radioactive source is a random variable. For a given range of energies (keV or spectrum channels), the number of gamma-ray counts registered by a detector per unit time and per unit amount of source material follows a Poisson distribution (Killian and Hartwell, 2000), $((\lambda)^k/k!)e^{-\lambda}$. The parameter λ primarily depends on the radioactive decay of the source, the characteristics of the detector material, with some contributions from environmental effects such as temperature effects on electronics and the position of the detector relative to the source. If β units of radioactive source material are present, the number of gamma-ray counts k follows

$$z \sim \frac{(\lambda\beta)^k}{k!}e^{-\lambda\beta}$$

where $\lambda\beta = Ez = E(z-\lambda\beta)^2$ is the mean value as well as the variance of the random variable. E stands for the expectation operator. $\beta = 0$ indicates that the source is absent. In the presence of multiple radioactive sources and the background radiation, the number of the gamma-ray counts registered at a detector is the sum of contributions from all sources and the background. For simplicity, the background may and will be considered as a radiation source.

In the presence of p radioactive sources, it is reasonable to assume that the contribution of individual source is statistically independent. Now, since the gamma-ray counts for each source is Poisson distributed,

$$\frac{(\lambda_i\beta_i)^k}{k!}e^{-\lambda_i\beta_i},\quad i = 1,2,\ldots,p$$

the total gamma-ray counts by the properties of the Poisson distribution is also Poisson distributed (Ash, 2008).

$$z \sim \frac{\left(\sum_{i=1}^{p}\lambda_i\beta_i\right)^k}{k!}e^{-\sum_{i=1}^{p}\lambda_i\beta_i}. \tag{1}$$

The above analysis applies to each range of energy (keV) of the spectra. Suppose a spectrum has N channels. Then, the gamma-ray counts is a random vector $Z = (z_1, z_2,\ldots, z_N)'$ and each z_i is the gamma-ray counts in a given channel that follows a Poisson distribution,

$$z_i \sim \frac{(\lambda_{i1}\beta_1 + \lambda_{i2}\beta_2 + \ldots + \lambda_{ip}\beta_p)^k}{k!}e^{-(\lambda_{i1}\beta_1+\lambda_{i2}\beta_2+\ldots+\lambda_{ip}\beta_p)},\quad i = 1,2,\ldots,N \tag{2}$$

where the vector $(\lambda_{1i}, \lambda_{2i},\ldots, \lambda_{Ni})^T$ is the spectrum of the ith radioactive source.

Define

$$\lambda = \begin{pmatrix}\lambda_{11} & \ldots & \lambda_{1p}\\ \vdots & \ddots & \cdots\\ \lambda_{N1} & \ldots & \lambda_{Np}\end{pmatrix},\quad \beta = \begin{pmatrix}\beta_1\\ \vdots\\ \beta_p\end{pmatrix} \tag{3}$$

where $(\lambda_{i1}\beta_1 + \lambda_{i2}\beta_2 +\ldots+ \lambda_{ip}\beta_p)$ is the mean gamma-ray counts per unit time received by the detector at the ith spectrum channel from all radioactive sources. For a little abuse of notation, the random gamma-ray counts vector Z follows

$$Z = \begin{pmatrix}z_1\\ z_2\\ \vdots\\ z_N\end{pmatrix} \sim \frac{(\lambda\beta)^k e^{-\lambda\beta}}{k!}$$

where $e^{-\lambda\beta}$ and $(\lambda\beta)^k$ are calculated component-wise. If we further assume that the number of gamma-ray counts at different channels are also independent, then the mean value and the variance of Z are given by

$$EZ = \lambda\beta = \begin{pmatrix}\lambda_{11} & \ldots & \lambda_{1p}\\ \vdots & \ddots & \cdots\\ \lambda_{N1} & \ldots & \lambda_{Np}\end{pmatrix}\begin{pmatrix}\beta_1\\ \vdots\\ \beta_p\end{pmatrix}$$

$$E(Z-\lambda\beta)(Z-\lambda\beta)^T = \begin{pmatrix}\lambda_{11}\beta_1+\ldots+\lambda_{1p}\beta_p & 0 & \ldots & 0\\ 0 & \lambda_{21}\beta_1+\ldots+\lambda_{2p}\beta_p & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \lambda_{N1}\beta_1+\ldots+\lambda_{Np}\beta_p\end{pmatrix}$$

Let Z_i, $i = 1, 2,\ldots, l$, be l independent observations of Z. Define the empirical average of Z by

$$\overline{Z} = \frac{1}{l}\sum_{i=1}^{l}Z_i$$

that satisfies

$$\overline{Z} = \lambda\beta + \underbrace{\frac{1}{l}\sum_{i=1}^{l}Z_i - EZ}_{\overline{V}} = \lambda\beta + \overline{V} \tag{4}$$

which is cast in a linear regression form and can be used to estimate the strength coefficients β where $\overline{Z}$ is available from measurements, the relative magnitude matrix λ is known for given radioactive sources and $\overline{V}$ is a random variable which can be considered as noise.

It is easy to verify

$$E\overline{V} = \frac{1}{l}\sum_{i=1}^{l}EZ^i - EZ = 0,$$

$$E\overline{V}\,\overline{V}^T = \frac{1}{l}\sum_{i=1}^{l}E\left(Z^i-\lambda\beta\right)\left(Z^i-\lambda\beta\right)^T$$

$$= \frac{1}{l}\begin{pmatrix}\lambda_{11}\beta_1+\ldots+\lambda_{1p}\beta_p & 0 & \ldots & 0\\ 0 & \lambda_{21}\beta_1+..+\lambda_{2p}\beta_p & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \lambda_{N1}\beta_1+\ldots+\lambda_{Np}\beta_p\end{pmatrix}$$

Notice that $\sum_{i=1}^{l}Z^i$ is also Poisson. If the gamma-ray counts are high, the Poisson $\sum_{i=1}^{l}Z^i$ is very close to a Gaussian distribution

(Ash, 2008). In other words, the noise $\overline{V}$ is close to a multivariate Gaussian of zero mean. Moreover, by the properties of the Gaussian, the components of $\overline{V}$ are independent if $\overline{V}$ are Gaussian.

3. Detection algorithms

In detection, Eq. (4) could be used to determine the unknown source strength coefficient β where $\overline{Z}$ is available from detector measurements and λ can be determined a priori. The purpose of detection is to find out if β_i's are zero or not. $\beta_i = 0$ indicates that the ith radioactive source or material is absent while $\beta_i > 0$ indicates the presence of the source. In general, the available background radiation information can also be included in λ and β. Although the components of the noise $\overline{V}$ are of zero mean and almost Gaussian and independent, their variance are not identical. In detection, it is always preferable to have iid noise. We note that if $\lambda\beta$ were available, by scaling $\overline{Z}$ as

$$\begin{pmatrix} \frac{1}{\sqrt{\lambda_{11}\beta_1+\ldots+\lambda_{1p}\beta_p}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\lambda_{N1}\beta_1+\ldots+\lambda_{Np}\beta_p}} \end{pmatrix}$$

$$\overline{Z} = \begin{pmatrix} \frac{1}{\sqrt{\lambda_{11}\beta_1+\ldots+\lambda_{1p}\beta_p}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\lambda_{N1}\beta_1+\ldots+\lambda_{Np}\beta_p}} \end{pmatrix} \lambda\beta + \begin{pmatrix} \frac{1}{\sqrt{\lambda_{11}\beta_1+\ldots+\lambda_{1p}\beta_p}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\lambda_{N1}\beta_1+\ldots+\lambda_{Np}\beta_p}} \end{pmatrix} \overline{V}$$

The noise

$$\begin{pmatrix} \frac{1}{\sqrt{\lambda_{11}\beta_1+\ldots+\lambda_{1p}\beta_p}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\lambda_{N1}\beta_1+\ldots+\lambda_{Np}\beta_p}} \end{pmatrix} \overline{V}$$

would have identical variance for each component. Though $\lambda\beta$ is not available, its estimate $\overline{Z}$ is however available. To reduce the effect of non-constant variance, Eq. (4) is scaled by

$$\underbrace{\begin{pmatrix} \frac{1}{\sqrt{\overline{z}_1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\overline{z}_N}} \end{pmatrix} \begin{pmatrix} \overline{z}_1 \\ \vdots \\ \overline{z}_N \end{pmatrix}}_{Y} = \underbrace{\begin{pmatrix} \frac{1}{\sqrt{\overline{z}_1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\overline{z}_N}} \end{pmatrix} \lambda\beta}_{X} + \underbrace{\begin{pmatrix} \frac{1}{\sqrt{\overline{z}_1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\overline{z}_N}} \end{pmatrix} \overline{V}}_{V}$$

or simply

$$Y = X\beta + V \tag{5}$$

where

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix},\ X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nn} \end{pmatrix} = \begin{pmatrix} x_1^T \\ \vdots \\ x_N^T \end{pmatrix},$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix},\ V = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}.$$

Y and X are available that are the scaled versions of the gamma-ray counts $\overline{Z}$ and λ respectively. V is noise of zero mean and almost iid Gaussian components. Eq. (5) is the basic equation for detection.

3.1. Traditional peak detection algorithms

Though there are several versions, the idea of the peak detection methods is to locate peaks, to fit the peak to a Gaussian function with some width, position and height and then to match the energy of the photo peak to a list of energies in a nuclide library (Killian and Hartwell, 2000) or alternatively to calculate the variance of the data and then to compare the expected variance (Jarman et al., 2003). The key is a peak location algorithm which can be an early version (Mariscotti, 1967) or a more recent algorithm (Killian and Hartwell, 2000) that has some advanced smoothing functions with varying smooth windows to combat data fluctuations due to noise. The choice of the smoothing window size is to balance the peak detection sensitivity and the sensitivity of the noise.

If signals are strong and high-resolution detectors can be used, gamma-ray signature recognition is straightforward by peak detection algorithms (Murray, 1998). However, in a number of applications in which the detectors cannot reasonably be expected to be close to the sources, the detectors generally have large volume and poor resolution. The signals from these detectors will be weak and difficult to separate from the background radiation or from the signatures of commonly used radioactive materials. In such cases, the peak detection algorithms do not work well.

To illustrate, consider a numerical simulation example which involves 11 isotopes as shown in Table 1 and the background radiation which consists primarily of trace amounts of radiation from local sources and cosmic rays that is considered to be the 12th radioactive source.

For a germanium type of detector capable of resolving spectra to 1024 channels and the given 12 isotopes as shown in Table 1, their mean gamma-ray counts are well documented which provides λ in Eq. (4) as shown in Fig. 1. In simulation, the coefficients β_{12} of the background and β_9 of I131 are set to be 1 and the coefficient β_1 for Pu239 is equal to 0.2. All other β_i's are zero. Simply put, the background and I131 are present and dominant. Pu239 is also present but weakly. The top diagram of Fig. 2 shows the mean gamma counts of the background + I131 per unit time for channels 1 to 1024 and the mean gamma counts of the background + I131 + 0.2Pu239. The addition of 0.2Pu239 is hardly visible. The bottom diagram of Fig. 2 is a closed-up one together with the mean gamma counts of Pu239 per unit time. The signal to noise (background) ratio is calculated by summing up the energy at all frequencies,

Table 1
Isotopes involved in testing.

Isotope	1	2	3	4	5	6	7	8	9	10	11	12
	Pu239	Ga67	Cs137	U235	K40	Na22	Ba133	Ce139	I131	Co57	Co60	Background

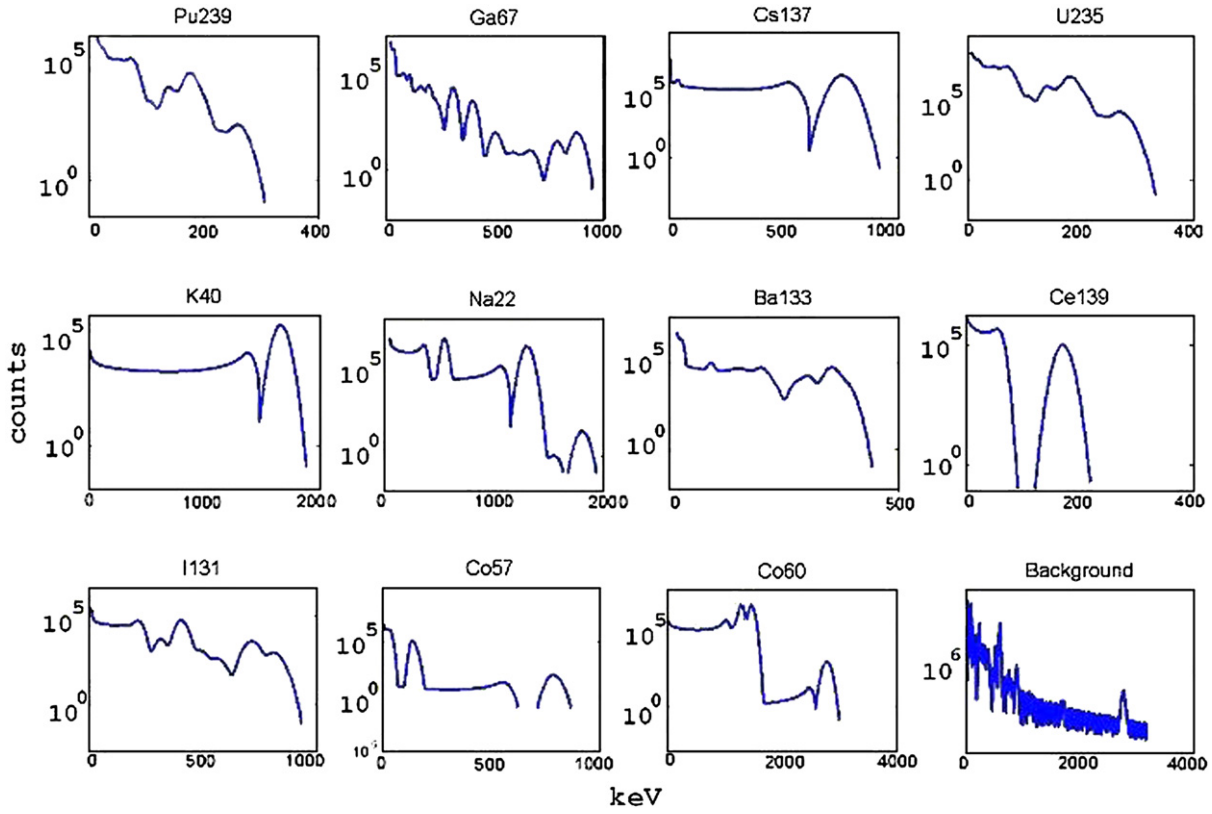

Fig. 1. Mean gamma-ray counts λ per unit time for each isotope.

$$10\cdot\log\frac{\sum_{i=1}^{1024}\{\alpha(0.2\mathrm{Pu239}(i)+\mathrm{I131}(i))\}}{\sum_{i=1}^{1024}\mathrm{Background}(i)}=-20\ \mathrm{dB},\quad \alpha=1.$$

For simulations in the paper, $\alpha = 1$ or $SNR = -20$ dB unless specified otherwise. It is obvious that all the signature peaks generated by Pu239 are buried in background or in I131 and are invisible which implies that traditional detection methods based on the peak detections of the gamma-ray counts will not work well. To test, we use an advanced peak detection algorithm developed at the University of Maryland (O'Haver, 2009) called findpeaks.m in MATLAB.

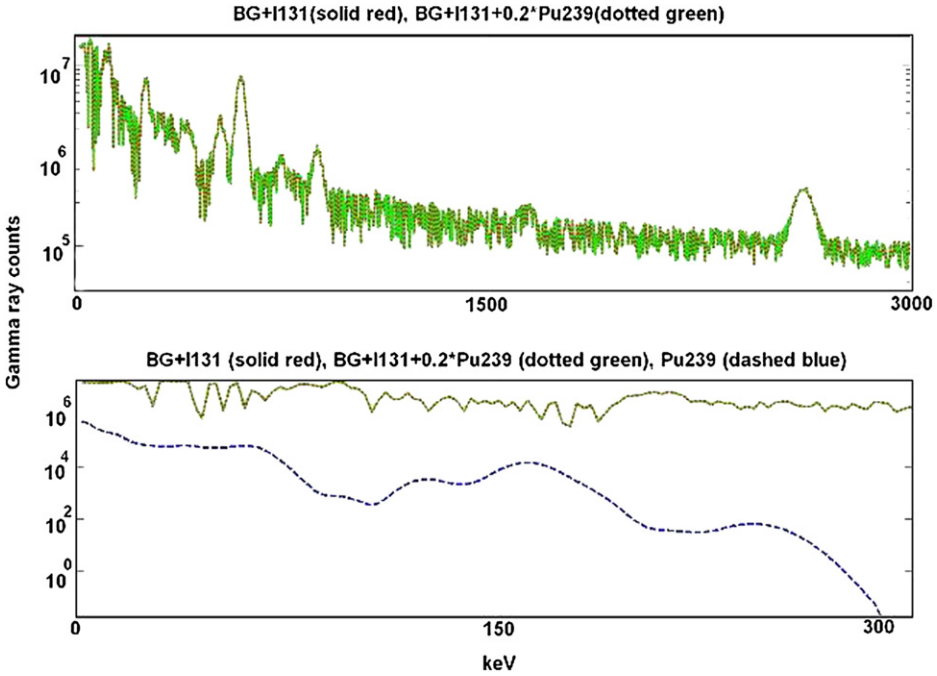

Fig. 2. Top: Gamma-ray counts of background + I131 (solid red) and background + I131 + 0.2Pu239 (dotted green). Bottom: Close-up with PU239 (dash–dash blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The peak detection results are shown in Figs. 3 and 4. For both the smoothing window sizes 3 and 7, the peaks detected with or without Pu239 remain the same. This implies that the traditional radionuclide identification algorithms based on peak detection would not work in this simulation example because they fail to detect signature peaks generated by Pu239.

3.2. Lasso algorithm

To overcome the problem of detecting invisible peaks, we present variable selection algorithms based on the model (5). Recall that our primary goal is not to estimate the magnitude of each component β_i, $i = 1, 2,\ldots, p$, but to find out if β_i is zero or not, referred to as the variable selection in the literature.

The Eq. (5) is in a standard linear equation form for the unknown β. Let $\hat{\beta}$ denote the estimate of β. It is well known that the least squares estimate

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T Y = \beta + \left(X^T X\right)^{-1} X^T V \tag{6}$$

is not appropriate for this purpose. A very popular and numerically efficient algorithm for variable selection is the celebrated Lasso (least absolute shrinkage and selection operator) algorithm (Zou et al., 2007) which jointly minimizes the fitting error and the L_1 penalty,

$$\min \|Y - X\hat{\beta}\|^2 + t\sum_{i=1}^{p} |\hat{\beta}_i| \tag{7}$$

The tuning parameter $t \geq 0$ controls the amount of penalty applied to the estimate $\hat{\beta}$. The tuning parameter $t = 0$ leads to the least squares and a very large t shrinks $\hat{\beta}$ to zero.

For the Lasso algorithms, the choice of the tuning parameter t is critical. In general, the choice of t is based on the well established AIC (Akaike Information Criterion (Shao, 1997)) and BIC (Bayesian Information Criterion (Shao, 1997)). To emphasize the dependence on t, we write the estimate $\hat{\beta}(t)$ derived from Eq. (7) for a given t. For each t and $\hat{\beta}(t)$, let

$$s(t) = \|Y - X\hat{\beta}(t)\|^2$$

and $q(t)$ be the number of non-zero components of $\hat{\beta}(t)$. If no component of $\hat{\beta}(t)$ is zero, $q(t) = p$, the dimension of β and $q(t) = 0$ if all the components of $\hat{\beta}(t)$ are zero. Define the optimal t as

$$t(\text{optimal}) = \arg\min_t \left\{ N \cdot \ln\left(\frac{s(t)}{N}\right) + \alpha q(t) \right\} \tag{8}$$

The optimal t according to the AIC or BIC is the one when $\alpha = 2$ or $\alpha = \ln(N)$, respectively (Shao, 1997). The Lasso(AIC) estimate is the estimate derived from Eq. (7) when t is chosen from Eq. (8) for $\alpha = 2$ while the Lasso(BIC) estimate is the one derived from Eq. (7) when t is chosen from Eq. (8) for $\alpha = \ln(N)$. Both the AIC and BIC estimates can be efficiently calculated (Zou et al., 2007).

To illustrate the performance of the Lasso(BIC) and the Lasso(AIC), we continue the same simulation study discussed previously where the mean gamma-ray counts provide λ in Eq. (4) as shown in Fig. 1 and further in Eq. (5) after the scaling.

The simulation is based on the model (5) with $l = 5$ and $N = 1024$. In simulation, the number of gamma-ray counts are generated according to Poisson distribution (2) for each isotope and each spectrum channel to have Z and $\overline{Z}$. Then, λ and $\overline{Z}$ are scaled by $\overline{Z}$ resulting in X and Y as in Eq. (5). V is the resulting random noise acting on the model (5). As before, the coefficients β_{12} of the background and β_9 of I131 are set to be 1 and the coefficient β_1 for Pu239 is equal to 0.2. All other β_i's are zero.

The Lasso(BIC) and the Lasso(AIC) are implemented in MATLAB with one modification. In our application, the values of the unknown β_i's are always non-negative. If the ith isotope is present, the corresponding β_i can be small but is positive. If the ith isotope is absent, the corresponding β_i is zero but cannot be negative. Thus, in numerical simulations, if $\hat{\beta}_i$ of the Lasso(BIC) or the Lasso(AIC) estimate is negative, we set $\hat{\beta}_i$ to zero and consider that the ith isotope is absent.

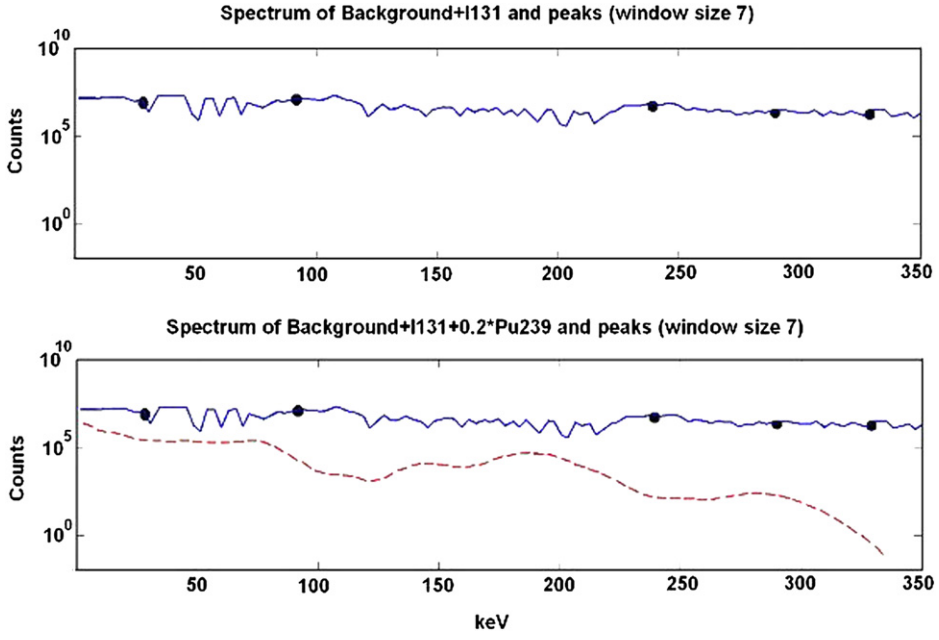

Fig. 3. Top: Peaks detected in circles (window size = 7 channels) of background + I131. Bottom: Peaks detected in circles (window size = 7 channels) of background + I131 + 0.2+Pu239 (solid blue) with Pu239 (dashed red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

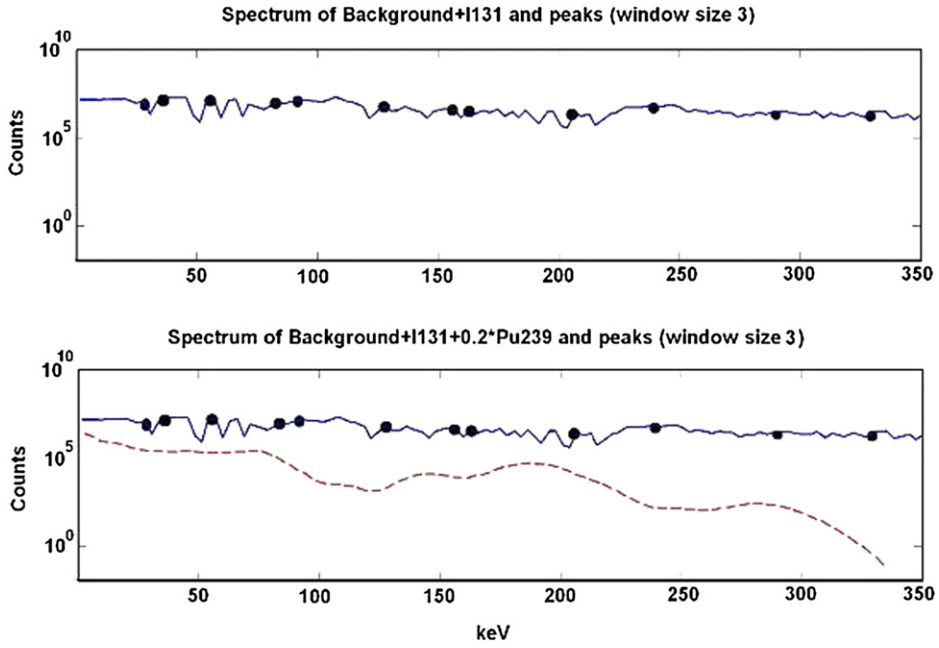

Fig. 4. Top: Peaks detected in circles (window size = 3 channels) of background + I131. Bottom: Peaks detected in circles (window size = 3 channels) of background + I131 + 0.2Pu239 (solid blue) with Pu239 (dashed red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To quantify the performance of the algorithm, we define two types of errors.

False negative: Isotope is present and the algorithm fails to find it.
False positive: Isotope is absent and the algorithm falsely identifies it.

The simulation consists of 10,000 Monte Carlo runs. Since for each run, I131 and Pu239 are always there and no other isotopes are present, the second row of Table 2 shows the total numbers of appearances of each isotope in 10,000 runs. The last two rows of Table 2 show the number of times each isotope was identified by the Lasso(BIC) and the Lasso(AIC), respectively. The corresponding errors are in the first two rows of Table 3.

From the simulation results, we can conclude that both the Lasso(BIC) and the Lasso(AIC) are very good in identifying the radioactive sources while their false positive rates are concerns.

To reduce the false positive errors, we discuss the performance limitation of the Lasso algorithm in the context of our application. Two questions are particularly interesting and important. First, for a given measurement data, does there always exist at least one tuning parameter t in Eq. (7) that provides the correct estimate $\hat{\beta}$? By correctness, we mean the Lasso estimates $\hat{\beta}_i \neq 0$ if $\beta_i \neq 0$ and $\hat{\beta}_i = 0$ if $\beta_i = 0$ for all i. Secondly, if such an optimal t exists, does the AIC or the BIC always finds such an optimal t? Since the analysis is similar, we focus on the AIC here.

The first question was studied in the literature (Leng et al., 2004) and the answer is no. In fact, the probability that no such t exists could be very high (Leng et al., 2004). To study the second question, consider a scalar equation

$$y(k) = \beta + v(k), \quad k = 1, 2, \ldots, N \tag{9}$$

where $v(k) = \xi(k) - \lambda$ and $\xi(k)$'s are independently Poisson distributed with mean and variance λ. In other words, $v(k)$'s are iid of zero mean and identical variance λ. Following the same procedures as in Zou et al. (2007), the Lasso estimate of Eq. (7) for a given t assumes the form of

$$\hat{\beta} = sign\left(\hat{\beta}^0\right)\left(|\hat{\beta}^0| - \frac{t}{2N}\right)^+$$

where $\hat{\beta}^0$ is the least squares estimate of β and $a^+ = a$ if $a > 0$ and $a^+ = 0$ if $a \leq 0$.

To find the optimal t based on the AIC and the corresponding Lasso estimate $\hat{\beta}$, we need to calculate two cases: (1) $|\hat{\beta}^0| \leq t/2N$ or $\hat{\beta} = 0$ and $q(t) = 0$ as in Eq. (8) which implies $\ln(s(t)/N) + \alpha(q(t)/N) = \ln \sum y(k)^2/N$, and (2) $|\hat{\beta}^0| > t/2N$ or $|\hat{\beta}| > 0$, $q(t) = 1$ and

$$\min\hat{\beta}\left\{\ln \sum \frac{(y(k) - \hat{\beta})^2}{N} + \frac{2}{N}\right\} = \ln \sum \frac{\left(y(k) - \hat{\beta}^0\right)^2}{N} + \frac{2}{N}.$$

Clearly, based on the AIC, the Lasso estimate $\hat{\beta} = 0$ if

$$\ln \sum \frac{y(k)^2}{N} < \ln \sum \frac{\left(y(k) - \hat{\beta}^0\right)^2}{N} + \frac{2}{N}$$

and $|\hat{\beta}| > 0$ if

$$\ln \sum \frac{y(k)^2}{N} > \ln \sum \frac{\left(y(k) - \hat{\beta}^0\right)^2}{N} + \frac{2}{N}.$$

Now, suppose the actual $\beta = 0$, $\lambda = 10$ and $N = 1024$ in Eq. (9). The probability that the Lasso estimate based on the AIC is not zero

$$\text{Prob}\{|\hat{\beta}| > 0\} = \text{Prob}\left\{\ln \sum \frac{y(k)^2}{N} > \ln \sum \frac{\left(y(k) - \hat{\beta}^0\right)^2}{N} + \frac{2}{N}\right\} \approx 0.16$$

Table 2
The numbers of presence and detections by the Lasso(BIC)(LB) and the Lasso(AIC)(LA) for each isotope in 10,000 Monte Carlo runs.

#	Pu239	Ga67	Cs137	U235	K40	Na22	Ba133	Ce139	I131	Co57	Co60
Present	10,000	0	0	0	0	0	0	0	10,000	0	0
Detected (LA)	10,000	6579	6728	9725	5607	6101	6467	5791	10,000	6113	4054
Detected (LB)	10,000	8467	7911	9792	6172	7347	8534	7340	10,000	7734	2768

can be computed. This demonstrates that even in the case of $\beta = 0$ or the nuclear material is absent and an optimal $t > 2N|\hat{\beta}^0|$ exists that produces a correct estimate $\hat{\beta} = 0$, the AIC gives rise to a tuning parameter t that results in an incorrect estimates $|\hat{\beta}| > 0$ with a substantial probability. This explains why the false positive error of the Lasso algorithm is so high.

The above analysis clearly illustrates the problem of the Lasso, i.e., the Lasso or similar types of variable selection algorithms including the LARS, the Forward Stepwise, the Backward Stepwise and others variable selection algorithms (Zou et al., 2007) alone are unable to solve the radionuclide detection problem reliably and accurately.

3.3. Subsampling

On one hand, we want to utilize the nice property of the Lasso that has very small false negative errors. On the other, we need to reduce or eliminate its large false positive errors. Observe that the problem with the false positive error rate is that though the ith isotope is absent, the Lasso(BIC) or the Lasso(AIC) estimate $\hat{\beta}_i$ is a random variable that could be positive for a particular experiment or a Poisson realization. Notice that individual classification methods are recently challenged by combined classification systems, which often show better performance. A motivation for ensembles is that a combination of decision made by individual classification produces powerful committee (Hastie et al., 2001). Based on the idea of majority voting, we combine the Lasso with an estimation of a tight interval in which the true but unknown β lies. To this end, the least squares (LS) estimate and the probability distribution associated with the LS estimate is useful. Although the LS estimate is not good for variable selection in general because it usually gives a small non-zero value even the true value is zero, its asymptotical distribution does provide a way to estimate an interval that the true β lies in.

Let the LS estimate $\hat{\beta}$ be defined in Eq. (6). Its mean value and variance are given by

$$E\hat{\beta} = \beta,\ \ E(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T = \frac{\sigma^2}{N}\left(\frac{1}{N}X^TX\right)^{-1} = \frac{\sigma^2}{N}\left(\frac{1}{N}\sum_{i=1}^{N}x_ix_i^T\right)^{-1}$$

If $N\to\infty$, the asymptotic distribution of $\hat{\beta}$ is given by Garatti and Bitmead (2010)

$$\sqrt{N}(\hat{\beta}-\beta)\sim N\left(0,\ \sigma^2\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}x_ix_i^T\right)^{-1}\right)$$

or for large N, $\hat{\beta}$ is asymptotically Gaussian distributed $N(\beta,\ (\sigma^2/N)(1/N\sum_{i=1}^{N}x_ix_i^T)^{-1})$ and each component $\hat{\beta}_i$ obeys

Table 3
Detection errors of the Lasso(BIC) and the Lasso(AIC).

	False negative	False positive
Lasso(AIC)	0 (0/20,000 = 0)	57,165 (57,165/90,000 = 0.6352)
Lasso(BIC)	0 (0/20,000 = 0)	66,065 (66,065/90,000 = 0.7341)

$$N\left(\beta_i,\ \frac{\sigma^2}{N}\left(\frac{1}{N}\sum_{i=1}^{N}x_ix_i^T\right)^{-1}_{i,i}\right) \tag{10}$$

where $((1/N)\sum_{i=1}^{N}x_ix_i^T)^{-1}$ are computable and X_{ii} is the iith element of the matrix X.

Let *std* stands for the standard deviation of the distribution (10). By the property of the Gaussian distribution, with probability 0.999 or higher, $\beta_i - 3\cdot std \le \hat{\beta}_i \le \beta_i + 3\cdot std$ or equivalently

$$\hat{\beta}_i - 3\cdot std \le \beta_i \le \hat{\beta}_i + 3\cdot std.$$

Let $d > 0$ be a small positive number. If

$$\hat{\beta}_i + 3std < d \tag{11}$$

we can say with a high confidence that $\beta_i \le 0$ or the ith isotope is absent independent of the results of the Lasso(BIC) or the Lasso(AIC). This would provide us a way to reduce the false positive rates.

The problem is that actual LS distribution $F_{\hat{\beta}}$ is unknown and the approximation of Eq. (10) is good only for large enough N. For our application, N is fixed and may not be large. In such a case, there is no error bound available to guarantee the approximation performance. We need to develop a different approach that possesses a bound for finite N. We propose a subsampling approach to find the LS distribution $F_{\hat{\beta}}$. Rewrite the Eq. (5) as

$$y_i = x_i^T\beta + v_i,\quad i = 1,\ldots,N$$

Let N_s and m be two positive numbers such that

$$0 < N_s < N,\ \ m = N - N_s + 1$$

Define

$$X(i) = \begin{pmatrix} x_i^T \\ \vdots \\ x_{i+N_s-1}^T \end{pmatrix},\ \ Y(i) = \begin{pmatrix} y_i \\ \vdots \\ y_{i+N_s-1} \end{pmatrix},\ \ V(i) = \begin{pmatrix} v_i \\ \vdots \\ v_{i+N_s-1} \end{pmatrix},\ \ i = 1,2,\ldots,m$$

Further, let $\hat{\beta}(i)$ be the LS estimate of β based on the data $X(i)$ and $Y(i)$ of length N_s,

$$\hat{\beta}(i) = \left(X^T(i)X(i)\right)^{-1}X^T(i)Y(i) = \beta + \left(\frac{1}{N_s}\sum_{j=i}^{i+N_s-1}x_ix_i^T\right)^{-1}\left(\frac{1}{N_s}\sum_{j=i}^{i+N_s-1}x_iv_i\right),\ \ i = 1,2,\ldots,m$$

Let $F_{\hat{\beta}_{N_s}}(b) = \text{Prob}\{\hat{\beta}(i) \le b\}$ be the probability distribution function of $\hat{\beta}(i)$, where the vector inequality is taken componentwise. The idea is to approximate $F_{\hat{\beta}}$ of $\hat{\beta}$ by $F_{\hat{\beta}_{N_s}}$ of $\hat{\beta}(i)$.

Let

$$1_{[\hat{\beta}(i)\le b]} = \begin{cases} 1, & \hat{\beta}(i) > b \\ 0, & \hat{\beta}(i) \le b \end{cases}$$

be the indicator function and define the empirical probability function $\hat{F}_{\hat{\beta}_{N_s}}(b)$ as

$$\hat{F}_{\hat{\beta}_{N_s}}(b) = \frac{1}{m}\sum_{i=1}^{m} 1_{[\hat{\beta}(i)\le b]}$$

Then, we have

Theorem 3.1 Assume the y_i's are stationary and $(X^T(i)X(i))^{-1}$ exists for each i. Then,

1. If the v_i's are iid, then, we have for each m,

$$E\left(\hat{F}_{\hat{\beta}_{N_s}}(b) - F_{\hat{\beta}_{N_s}}(b)\right)^2 \le \frac{12}{m}$$

2. If the v_i's are generated by a stable linear filter derived by iid noise, then we have

$$E\left(\hat{F}_{\hat{\beta}_{N_s}}(b) - F_{\hat{\beta}_{N_s}}(b)\right)^2 \le 12\frac{N_s-1}{N-N_s}\left(2-\frac{N_s}{N-N_s}\right) + \frac{24}{N-N_s}\cdot\frac{\rho}{1-\rho}$$

for some $0<\rho<1$.

Proof: The second part of the theorem follows directly from Garatti and Bitmead (2010). For the first part, $F_{\hat{\beta}_{N_s}}(b) = E1_{[\hat{\beta}(i)\le b]}$. Observe $W_i = 1_{[\hat{\beta}(i)\le b]} - F_{\hat{\beta}_{N_s}}(b)$ is zero mean and stationary because y_i and v_i are stationary. Following (Garatti and Bitmead, 2010), we have

$$E\left(\hat{F}_{\hat{\beta}_{N_s}}(b) - F_{\hat{\beta}_{N_s}}(b)\right)^2 = E\left(\frac{1}{m}\sum_{i=1}^{m} W_i\right)^2 = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} W_iW_j = \frac{1}{m^2}\sum_{\tau=-m}^{m}(m-|\tau|)\gamma(\tau)$$

where $\gamma(\tau) = EW_iW_{i+\tau}$. Notice v_i is iid and its strong mixing coefficient satisfies

$$\alpha(\tau) = \sup_{A,B}\{|\text{Prob}\{AB\} - \text{Prob}\{A\}\text{Prob}\{B\}| | A\in A_0,\ B\in A^\tau\} = \begin{cases}\le 1 & \tau = 0\\ 0, & \tau>0\end{cases}$$

where A_0 and A^τ are the σ-algebras generated by v_i, $i\le 0$ and v_τ, $\tau\ge 0$ respectively. Now, it follows from Garatti and Bitmead (2010), $|\gamma(\tau)|\le 12\alpha(|\tau|)$. Thus,

$$E\left(\hat{F}_{\hat{\beta}_{N_s}}(b) - F_{\hat{\beta}_{N_s}}(b)\right)^2 = \frac{1}{m^2}\sum_{\tau=-m}^{m}(m-|\tau|)\gamma(\tau) \le \frac{12}{m}.$$

To demonstrate the improvement of the subsampling approach over the asymptotic approximation, the actual distribution of $\hat{\beta}_1$ by 100 Monte Carlo simulations and approximate distributions by asymptotic approximation and subsampling respectively are shown in Fig. 5. Clearly, the distribution by subsampling is closer to the actual one than the one by asymptotic approximation.

Remarks:

- The error bound given by the theorem is a non-asymptotic bound which provides an error quantification for modest N which is more useful than the asymptotic bound.

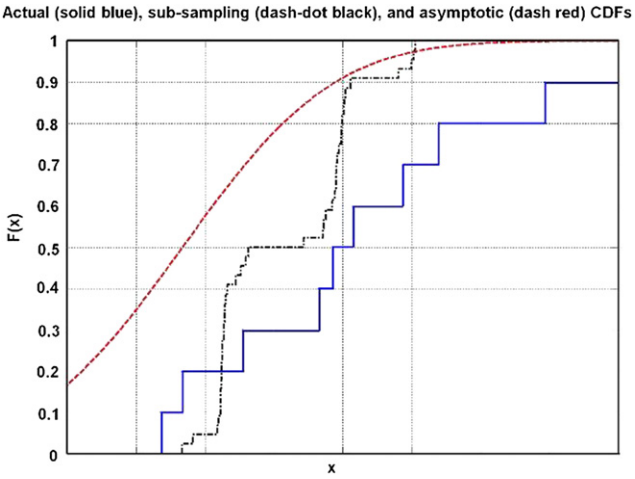

Fig. 5. Actual distribution (solid blue) and approximate distributions by asymptotic (dash–dash red) and subsampling (dash–dot black) approaches. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- What we are really interested is the LS estimate distribution $F_{\hat{\beta}}$ of $\hat{\beta}$ that is the estimate by taking all the data $i = 1,\ldots, N$. On the other hand, $F_{\hat{\beta}_{N_s}}$ is the LS distribution of $\hat{\beta}(i)$ by taking the data length $N_S < N$. Further, $F_{\hat{\beta}_{N_s}}$ is unknown and is approximated by $\hat{F}_{\hat{\beta}_{N_s}}$. Intuitively, when N_s is close to N, $F_{\hat{\beta}_{N_s}}$ is close to $F_{\hat{\beta}}$. However, N_s close to N implies $m = N - N_s + 1$ is small and the error between $F_{\hat{\beta}_{N_s}}$ and its estimate $\hat{F}_{\hat{\beta}_{N_s}}$ is large. N_s and m have to be adjusted carefully.

We can now summarize the detection algorithm by combining the Lasso(BIC) or the Lasso(AIC) and subsampling techniques.

Detection algorithm: Given a detector and possible radioactive sources, construct the matrix λ as in Eq. (4).

Step 1: Observed the number of gamma-ray counts and construct the Eq. (5).
Step 2: Apple the Lasso(BIC) or the Lasso(AIC) as discussed previously. Set $\hat{\beta}_i$ to zero if $\hat{\beta}_i < 0$.
Step 3: Set the threshold $d > 0$ and m, N_s. Determine the sampled distribution $F_{\hat{\beta}_{N_s}}$ for each $\hat{\beta}_i$ using the subsampling technique. Calculate the sampled mean value $\hat{\mu}_i$ and the sampled standard deviation $\hat{\sigma}_i$ from the sampled distribution $F_{\hat{\beta}_{N_s}}$ for each $\hat{\beta}_i$.
Step 4: For both the Lasso(BIC) and the Lasso(AIC), if $\hat{\beta}_i > 0$ and $(\hat{\mu}_i + 3\hat{\sigma}_i) \ge d$, we say the ith isotope is present. Otherwise, the ith isotope is absent.

Table 4 shows the 10,000 Monte Carlo simulation results of the above detection algorithm with $d = 0.05$, $m = 44$ and $N_s = N - m = 980$ at $SNR = -20$ dB and Table 5 shows the results at different SNRs. By sacrificing the false negative rates slightly, the false positive rates have been drastically reduced from 0.6352 to 0.00078 for the Lasso(AIC) and from 0.7341 to 0.00099 for the Lasso(BIC) respectively. At the same time the false negative rates remain very small at 0.0034 for both the Lasso(AIC) and the Lasso(BIC). Clearly, the combined algorithm of the Lasso and the subsampling outperforms the both the Lasso(AIC) and the Lasso(BIC) significantly. An interesting question is how the performance of the combined algorithm is compared to the algorithm that is based on the subsampling only. To this end, the false negative and false positive errors of the subsampling algorithm from the same simulation with the

Table 4
The numbers of presence and detections of each isotope by the Lasso(BIC)(LB) and the Lasso(AIC)(LA) together with the subsampling at $SNR = -20$ dB.

#	Pu239	Ga67	Cs137	U235	K40	Na22	Ba133	Ce139	I131	Co57	Co60
Present	10,000	0	0	0	0	0	0	0	10,000	0	0
Detected (LA)	9932	68	0	0	0	0	2	0	10,000	0	0
Detected (LB)	9932	87	0	0	0	0	2	0	10,000	0	0

Table 5
Detection errors of the combined algorithms and the subsampling at different SNRs. LAS = Lasso(AIC)+Subsampling, LBS = Lasso(BIC)+Subsampling and Sub = Subsampling only.

	False negative rate					False positive rate				
	−15 dB	−18 dB	−20 dB	−22 dB	−24 dB	−15 dB	−18 dB	−20 dB	−22 dB	−24 dB
LAS	0	0	0.0034	0.0118	0.325	0.0007	0.0007	0.00078	0.0010	0.0078
LBS	0	0	0.0034	0.0118	0.325	0.0007	0.00088	0.00098	0.0012	0.0078
Sub	0	0	0.0043	0.0136	0.403	0.0012	0.00099	0.0011	0.0024	0.0091

same threshold d are shown in the last row of Table 5. The combined algorithm improves the false negative error of the subsampling algorithm by approximately (0.0043/0.0034=)20% and the false positive error by approximately (0.0011/0.00078=)40%.

4. Concluding remarks

In the paper, a model for detection of radionuclides is proposed and a corresponding two stage algorithm is developed based on the model. The model and the algorithm are particularly useful when the signals are weak and the traditional detection algorithms based on peak search fail. From the numerical simulations, the proposed algorithm shows a promising result with small false positive and false negative errors. The choice of the threshold d is however ad hoc. How to find such a threshold reliably is a direction of further research.

Acknowledgments

This work was supported in part by DoE grant DE-FG52-09NA29364.

References

Anderson, K., Jarman, K., Mann, M., Pfund, D., Runkl, M., 2008. Discriminating nuclear threats from benign sources in gamma-ray spectra using a spectral comparison ratio method. Journal of Radioanalytical and Nuclear Chemistry 276, 713–718.
Ash, R., 2008. Basic Probability Theory. Dover Books.
August, R., Whitlock, R., 2005. HELGA II: Autonomous passive detection of nuclear weapons materials. 2005 NRL Review. Naval Research Lab.
Fetter, S., Cochran, T., Grodzins, L., Lynch, H., Zucker, M., 1990. Gamma ray measurements of a soviet cruise missile warhead. Science 248, 828–834.
Garatti, S., Bitmead, R., 2010. On resampling and uncertainty estimation in linear system identification. Automatica 46 (5), 785–795.
Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer Verlag, New York.
Jarman, K., Daly, D., Anderson, K., Wahl, K., 2003. A new approach to automated peak detection. Chemometrics and Intelligent Laboratory Systems 69, 61–76.
Killian, E., Hartwell, J., 2000. PCGAP: Users guide and algorithm description. Idaho Natl. Engineering and Environmental Lab. Bechtel BWXT Idaho, LLC. INEEL/EXT-2000–00908.
Leng, C., Lin, Y., Wahba, G., 2004. A note on the lasso and related procedures in model selection. Statistica Sinica 16, 1273–1284.
Mariscotti, M., 1967. A method for automatic identification of peaks in the presence of background and its application to spectrum analysis. Nuclear Instruments and Methods 50, 309–320.
Moss, J., Donald, G., Lee, S., Jehanne, S., Keister, B., 2002. Report on the workshop on the role of the nuclear physics research community in combating terrorism. Washington, D.C., 11â"12 July 2002. U.S. Dep. of Energy, Washington, D.C. 20585-1290.
Murray, W., 1998. GN-4, a lightweight, battery-operated gamma-ray and neutron detection system for in situ nuclear material surveys. Los Alamos Natl. Lab. document LA-UR-97-2735.
O'Haver, T., 2009. Peak Finding and Measurement, Version 2. http://terpconnect.umd.edu/toh/spectrum/PeakFindingandMeasurement.htm.
Shao, J., 1997. An asymptotic theory for linear model selection. Statistica Sinica 7, 221–264.
Wei, W., Du, Q., Younan, N., 2010. Particle swarm optimization based spectral transformation for radioactive material detection and classification. In: IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA), Issue Date: 6–8 Sept. 2010, pp. 1–6.
Weller, H., Ahmed, M., Norum, B., McNabb, D., Litvinenko, V., 2006. Detection and isotopic identification of nuclear materials using a monoâ˜energetic gâ˜ray beam. Triangle Universities Nucl. Lab.
Zou, H., Hastie, T., Tibshirani, R., 2007. On the degrees of freedom of the lasso. The Annals of Statistics 35, 2173–2192.